

AZ

1/5/1 (Item 1 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2000 JPO & JAPIO. All rts. reserv.

04637465 \*\*Image available\*\*  
DOCUMENT PROCESSOR



PUB. NO.: 06-309365 JP 6309365 A]  
PUBLISHED: November 04, 1994 (19941104)  
INVENTOR(s): WATANABE MIKI  
KOMATSUBARA HIROFUMI  
APPLICANT(s): FUJI XEROX CO LTD [359761] (A Japanese Company or  
Corporation), JP (Japan)  
APPL. NO.: 05-115400 [JP 93115400]  
FILED: April 20, 1993 (19930420)  
INTL CLASS: [5] G06F-015/40; G06F-012/00; G06F-015/20  
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications); 45.2  
(INFORMATION PROCESSING -- Memory Units)

MAILED  
APR 17 2000  
Group 2790

#### ABSTRACT

PURPOSE: To provide the document processor which can easily generate retrieval information for retrieving a document or document element.

CONSTITUTION: When the document is registered in a document storage means 11, an index detecting means 121 generates the information for retrieval on the basis of an index and registers the information for retrieval even in a retrieval information storage means. For the generation of the information for retrieval, the index is detected first in the document. A retrieval information generating means 122 extracts the index character string corresponding to the position of the detected index, and generates and stores the retrieval information showing the correspondence relation between the index character string and the document containing it in a retrieval information storage means 13.

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-309365

(43)公開日 平成6年(1994)11月4日

(51)Int.Cl.<sup>5</sup>

G 0 6 F 15/40

12/00

15/20

識別記号

5 0 0 T

5 2 0 A

5 7 0 N

庁内整理番号

9194-51.

8944-5B

7315-5L

F I

技術表示箇所

審査請求 未請求 請求項の数 6 F D (全 27 頁)

(21)出願番号

特願平5-115400

(22)出願日

平成5年(1993)4月20日

(71)出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂三丁目3番5号

(72)発明者 渡辺 美樹

神奈川県川崎市高津区坂戸3丁目2番1号

K S P R&amp;Dビジネスパークビル 富

士ゼロックス株式会社内

(72)発明者 小松原 弘文

神奈川県川崎市高津区坂戸3丁目2番1号

K S P R&amp;Dビジネスパークビル 富

士ゼロックス株式会社内

(74)代理人 弁理士 岩上 昇一 (外3名)

(54)【発明の名称】 文書処理装置

(57)【要約】

【目的】 文書あるいは文書要素を検索するための検索用情報を容易に作成することができる文書処理装置を提供すること。

【構成】 見出し検出手段(121)は、文書を文書記憶手段(11)に登録する際に、検索用の情報を見出しを基に作成し、その検索用情報を検索用情報記憶手段にも登録する。検索用情報を作成するために、まず、文書から見出しを検出する。検索用情報生成手段(122)は検出した見出しの位置に対応する見出し文字列を抽出すると共に、その見出し文字列とそれを含む文書との対応関係を表す検索用情報を生成し、検索用情報記憶手段(13)へ記憶する。

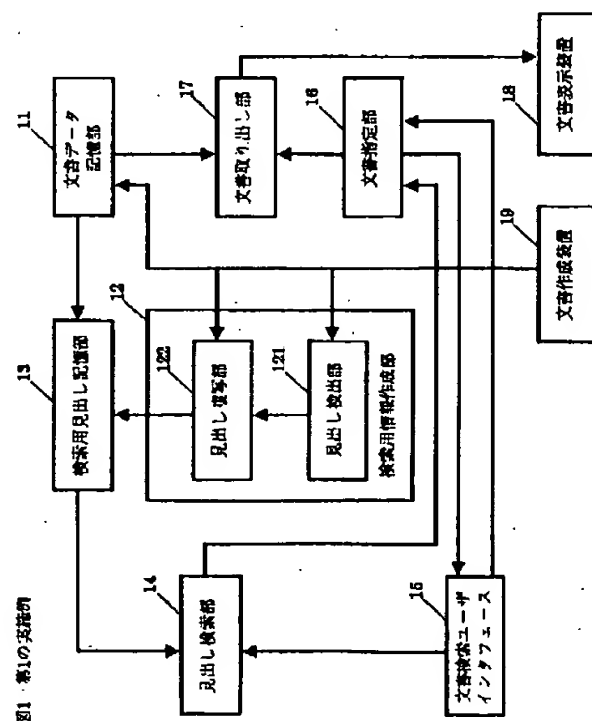


図1 第1の実施例

## 【特許請求の範囲】

【請求項1】 文書を記憶する文書記憶手段と、  
前記文書記憶手段に記憶される文書から見出しを表す文字列を検出する見出し検出手段と、  
前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された文書との対応関係を表す検索用情報を生成する検索用情報生成手段と、  
前記検索用情報生成手段により生成された検索用情報を記憶する検索用情報記憶手段とを設けたことを特徴とする文書処理装置。

【請求項2】 文書要素から構成される文書を記憶する文書記憶手段と、  
文書要素の属性を検索条件として指定して、文書記憶手段の文書から、検索条件を満たす文書要素を検索する文書要素検索手段とを設けたことを特徴とする文書処理装置。

【請求項3】 前記文書要素検索手段により取り出された文書要素を所定の形式で組み合わせて一つの文書を作成する作成手段を有することを特徴とする請求項2記載の文書処理装置。

【請求項4】 前記文書記憶手段に記憶される文書から見出しを表す文字列を検出する見出し検出手段と、  
前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された各文書の文書要素との対応関係を表す検索用情報を生成する検索用情報生成手段と、  
前記検索用情報生成手段により生成された検索用情報を記憶する検索用情報記憶手段とを設け、  
前記文書要素検索手段は、前記検索用情報記憶手段に記憶された検索用情報を用いて、指定の検索条件を満たす文書要素を取り出すことを特徴とする請求項3または請求項4記載の文書処理装置。

【請求項5】 文書記憶手段に記憶された文書を編集する文書編集手段と、  
前記文書編集手段により追加、変更、削除された見出しを抽出、記録する見出し編集記録手段と、  
前記見出し編集記録手段により記録された見出しの追加、変更、削除を検索用情報記憶手段に反映させる検索用情報更新手段とを設けたことを特徴とする請求項1または請求項4記載の文書処理装置。

【請求項6】 文書要素を記憶する文書要素記憶手段と、  
前記文書要素記憶手段に記憶される文書要素から見出しを表す文字列を検出する見出し検出手段と、  
前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された文書要素との対応関係を表す検索用情報を生成する検索用情報生成手段と、  
前記検索用情報生成手段により生成された検索用情報を記憶する検索用情報記憶手段とを設けたことを特徴とする文書処理装置。

【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、文書の登録および検索を容易にした文書処理装置に関する。

## 【0002】

【従来の技術】 大量の文書を登録し、必要に応じて登録された文書を検索し、文書の作成等に利用することのできる文書処理システムにおいて、目的の文書を速やかに検索するためにキーワードを各文書に対応して付けることが一般的に行われている。このキーワード検索を行う文書処理システムに文書を登録するにはその文書にふさわしいキーワードを付加することが必要であり、その文書の内容を変更したときには更新の必要がある。そのため文書の登録や更新をする作業が煩雑になる。この作業を軽減するための技術として特開平3-156678号公報記載の技術がある。これは登録すべき文書の編集中に文字列を指定することにより、その文字列を検索用の情報すなわちキーワードとして簡単に登録できるようにしたものである。

【0003】 文書を作成、編集する際に、既存の文書の一部を再利用することにより文書の作成効率を向上させようとするのが広く行われている。このような再利用を促進するために、文書からその内容の一部を文書部品として切り出して登録することが行われている。これは、例えば特開平3-157753号公報、特開平2-148250号公報、特開平3-8087号公報等に表示されている。しかし、これらの技術では、文書部品を再利用するためには、予め文書部品ごとにデータベースに登録する必要がある。

【0004】 このように文書から取り出された文書部品が大量になると、目的の文書部品を見つけ出すことが困難となる。大量の文書から目的の文書を見つけ出すための前述の特開平3-156678号公報のような、文書にキーワードを付加する技術を文書部品の検索に適用することが可能である。

【0005】 また、従来、文書検索手法として、文書名、著者名、キーワードなどの文書属性による検索が多用されている。文書属性検索では、所望する文書を検索しづらいことから、文書内容の全域を検索対象とする全文検索の手法が注目されている。全文検索は、検索対象となるデータが多くなることから、性能向上を行うために検索対象となる文書を予めスキャンし、検索用のインデックスの作成を行ったりしている。

【0006】 一方、属性検索を行い、結果が複数になった場合にどの文書が所望の文書であるかを判別するために文書の一部を表示させるという方式（特開平3-92961号公報）がある。この方式では、属性検索で該当した文書を検索の途中に該当する度に文書毎に文書の一部を表示するようになっている。

## 【0007】

【発明が解決しようとする課題】 文書の登録時にキーワ

## 3

ードを付ける従来技術（例えば、前記特開平3-156678号公報）ではキーワードとなる文字列を文書中から取り出し、それらを一つの検索用ファイルにまとめる方法を示しているが、キーワードとして登録する文字列を明示的に指定しなければならないため、長大な文書の登録の際は登録の作業が煩雑になるという問題がある。また、キーワードを登録すると、キーワードとして指定された文字列が通常の文書編集処理によって変更されたり削除された場合にその変更が検索データに反映されず、キーワードの登録が正確に行われないという問題がある。

【0008】また、前述の再利用したい文書部品だけを選択的に格納して、効率的に再利用を行なおうとする前述の従来技術（特開平2-148250号公報や特開平3-8087号公報等）では、文書部品を再利用するためには、予め文書部品ごとにデータベースに登録する必要がある。これは一々文書部品を個別に登録するという手間と再利用されるかどうかということを登録時に判断する必要があり、登録作業が面倒であるという問題がある。

【0009】また、属性検索を行い、結果が複数になった場合にどの文書が所望の文書かを判別するために文書の一部を表示させるという方式（特開平3-92961号公報）においては、文書毎に文書の一部が表示されるので、所望の文書かどうか判断するのに手間が掛かるとい問題、並びに、文書の表示すべき部分を指定できないという問題がある。

【0010】そこで、本発明は、文書あるいは文書要素を検索するための検索用情報を容易に作成することができる文書処理装置を提供することを課題とする。

【0011】また、本発明は登録文書または登録文書要素に変更が施されても、その変更が検索用情報に反映され、検索用情報を常に正確に保つことのできる文書処理装置を提供することを課題とする。

【0012】また、本発明は検索により複数の文書または文書要素が得られた時に、その中から所望のものを速やかに見つけ出すことのできる文書処理装置を提供することを課題とする。

【0013】

【課題を解決するための手段】本発明の文書処理装置は、文書を記憶する文書記憶手段（図1の11、図10の101）と、前記文書記憶手段に記憶される文書から章、節、図表等の見出しを表す文字列を検出する見出し検出手段（図1の121、図10の1021）と、前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された文書との対応関係を表す検索用情報を生成する検索用情報生成手段（図1の122、図10の1022）と、前記検索用情報生成手段により生成された検索用情報を記憶する検索用情報記憶手段（図1の13、図10の103）とを具備する。

## 4

【0014】また、本発明（請求項2）は、上記構成の文書処理装置において、文書記憶手段に記憶された文書を編集する文書編集手段（図10の108）と、前記文書編集手段により追加、変更、削除された見出しを抽出、記録する見出し編集記録手段（図10の109）と、前記見出し編集記録手段により記録された見出しの追加、変更、削除等の編集事項を検索用情報記憶手段に反映させる検索用情報更新手段（図10の1023）とを設けたことを特徴とする。

10 【0015】本発明（請求項3）は、文書要素から構成される文書を記憶する文書記憶手段（図17の171、図22の221）と、文書要素の属性を検索条件として指定して、文書記憶手段の文書から、検索条件を満たす文書要素を検索する文書要素検索手段（図17の174、図22の223）とを設けた文書処理装置である。

【0016】本発明（請求項4）は、前記発明（請求項3）の文書処理装置において、前記文書要素検索手段により取り出された文書要素を所定の形式で組み合わせる一つの文書を作成する作成手段（図22の224）を設けたものである。

20 【0017】本発明（請求項5）は、前記発明（請求項3）の文書処理装置において、前記文書記憶手段に記憶される文書から見出しを表す文字列を検出する見出し検出手段（図17の1721）と、前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された各文書の文書要素との対応関係を表す文書要素検索用情報を生成する検索用情報生成手段（図17の1722）と、前記検索用情報生成手段により生成された文書要素検索用情報を記憶する検索用情報記憶手段（図17の173）とを設け、前記文書要素検索手段（図17の174）は、前記検索用情報記憶手段に記憶された検索用情報を用いて、指定の検索条件を満たす文書要素を取り出すように構成したものである。

30 【0018】本発明（請求項6）は、文書要素を記憶する文書要素記憶手段（図7の71）と、前記文書要素記憶手段に記憶される文書要素から見出しを表す文字列を検出する見出し検出手段（図7の721）と、前記見出し検出手段により検出された文字列と前記文書記憶手段により記憶された文書要素との対応関係を表す文書要素検索用情報を生成する検索用情報生成手段（図7の722）と、前記検索用情報生成手段により生成された文書要素検索用情報を記憶する検索用情報記憶手段（図7の73）とを設けた文書処理装置である。

【0019】

40 【作用】本発明（請求項1）の文書処理装置において、見出し検出手段（121）は、文書を文書記憶手段（11）に登録する際に、検索用の情報を作成し、その検索用情報を検索用情報記憶手段にも登録する。検索用情報を作成するために、まず、文書から見出しを検出する。50 文書記憶手段（11）に格納される対象の文書として構

造化文書を対象とする場合には、その構造により表題、見出し、図形、本文などを区別する情報を有しているため、見出し検出手段はそれらを調べることにより見出しの位置を検出する。構造化文書ではない一般の文書を対象とする場合には、文書を解析し、見出しの特徴を有する部分、例えば数字で始まる行や前後の行と異なるフォントを用いた行等、を検出する。検索用情報生成手段

(122)は検出した見出しの位置に対応する見出し文字列を抽出すると共に、その見出し文字列とそれを含む文書との対応関係を表す検索用情報を生成し、検索用情報記憶手段(13)へ記憶する。本発明によれば、文書に含まれる見出しを見出し検出手段により抽出して、検索用情報を作成し、従来のキーワードの代わりに登録するようにしたので、キーワードを選定する煩雑な作業を要することなく容易に検索対象文書の登録が可能になる。

【0020】本発明(請求項2)において、文書記憶手段(図17の171、図22の221)には、章、節、図、表、イメージのような文書要素から構成される文書(図18、図19)が記憶されている。文書要素検索手段(図17の174、図22の223)は、文書要素の属性が検索条件として与えられると、文書記憶手段の文書を検索し、検索条件を満たす文書要素を取り出す。本発明によれば、文書要素を選択的に格納するのではなく、文書そのものを格納し、検索は文書要素単位で取り出すようにしたことにより、既存文書を構成する全ての文書要素を再利用の対象とすることができる。さらに文書として格納しているので、格納時に、いちいち文書要素の指定をする手間が省くことができ、また、文書の編集のみを行えば文書要素を検索用に別途編集する必要がない。

【0021】前記文書要素を単位として検索することのできる発明(請求項2)において作成手段(図22の224)を設けた発明(請求項3)は、前記文書要素検索手段により取り出された複数の文書要素を所定の形式で組み合わせて一つの文書を作成する。例えば、ある論文の集まりからそのアブストラクトのみを抽出し、そのアブストラクトと文書名を組みにした段落で構成される一つの仮想文書を作成できる(例えば図27)。この例の場合、ユーザにその仮想文書を提示することにより、所望の論文を容易に探すことができる。本発明によれば、既存文書の所望の部分抽出して一つの別の文書としてユーザに提示することができるので、ユーザは提示された文書を見て必要な文書の選択をすることができ、また、文書の要素を格納時の目的とは別の複数の目的に再利用することができる。

【0022】文書全体を登録した文書からの文書要素の検索は、その文書が構造化文書である場合には、その文書構造のルートを追うことにより容易におこなうことができるが、前述の見出しを抽出して検索用情報を生成

し、この検索用情報により検索するようにした場合(請求項4)には、文書要素の登録が容易になると共に、より高速の検索が可能となる。

【0023】文書記憶手段に記憶された文書または文書部品に対して文書編集手段により編集可能に構成した本発明(請求項5)の文書処理装置においては、文書編集手段による見出しの追加、変更、削除などを見出し編集記録手段により記録する。検索用情報更新手段は、この記録に基づいて検索用情報を更新する。これにより、編集の結果が検索用情報に反映され、文書の検索を常に正確になるように維持できる。

【0024】本発明(請求項6)は、文書要素記憶手段(71)に文書要素を単位として登録するものにおいて、見出しを抽出して検索用情報を生成するものである。即ち、見出し検出手段(721)は、文書要素を文書要素記憶手段に登録する際に、検索用情報を作成し、その検索用情報を検索用情報記憶手段(73)にも登録する。検索用情報を作成するために、まず、文書要素から見出しを検出する。文書要素記憶手段に格納される対象の文書要素として構造化文書に適用される形式を持つものを対象とする場合には、その構造により文書要素が表題、見出し、図形、本文などを区別する情報を有しているので、見出し検出手段はそれらを調べることにより見出しを含む文書要素を検出する。構造化文書ではない一般の文書を対象とする場合には、文書部品を解析し、見出しの特徴を有する部分、例えば数字で始まる行や前後の行と異なるフォントを用いた行等、を検出する。検索用情報生成手段(722)は検出した見出しを含む文書要素にある見出し文字列を抽出すると共に、その見出し文字列とそれを含む文書要素との対応関係を表す検索用情報を生成し、検索用情報記憶手段へ記憶する。本発明によれば、文書要素に含まれる見出しを見出し検出手段により抽出して、検索用情報を作成し、従来のキーワードの代わりに登録するようにしたので、キーワードを選定する煩雑な作業を要することなく容易に検索対象文書要素の登録が可能になる。

【0025】

【実施例】

(第1の実施例)図1は本発明の第1の実施例の文書処理装置の構成を示すブロック図である。この装置は、図1に示すように、文書データ記憶部11、検索用情報作成部12、検索用見出し記憶部13、見出し検索部14、文書検索ユーザインタフェース15、文書指定部16、文書取り出し部17、文書表示装置18および文書作成装置19からなっている。

【0026】文書データ記憶部11は文書作成装置19により作成された文書を記憶するために用いられる。

【0027】検索用情報作成部12は、文書作成装置19からの文書データを文書データ記憶部11に記憶する際に、文書データに含まれる見出しから検索用情報を作

成するものであり、文書作成装置19により作成された文書データを解析し、章、節、図表の見出しとなる部分を検出する見出し検出部121と、検出した見出しの文字列を複写し文書データとの対応を付けて出力する見出し複写部122とを有している。

【0028】検索用見出し記憶部13は、検索用情報作成部12により作成された、見出しとなる文字列と文書データ記憶部11に記憶される文書データとの対応付けを表す検索用情報を記憶するためのものである。

【0029】見出し検索部14は文書を検索する際に用いられ、文書検索ユーザインタフェース15により与えられた文字列と同じまたはそれを含む見出しを検索用見出し記憶部13から検索し、該当する見出しを含む文書の一覧を作成するためのものである。

【0030】文書検索ユーザインタフェース15は、見出し検索部14に渡す文字列をユーザが入力したり、与えられた文書の一覧を表示しユーザから目的とする文書を文書指定部16により指定させるための入出力を司るために用いられる。

【0031】文書指定部16は見出し検索部14より作成された文書の一覧を文書検索ユーザインタフェース15により表示し、取り出すべき文書をユーザに選択させるためのものである。

【0032】文書取り出し部14は、文書指定部16により指定された文書データを文書データ記憶部11から取り出すためのものである。

【0033】文書表示装置18は文書取り出し部17により取り出された文書を表示させるために用いられる。

【0034】文書作成装置19は文書データを作成するために用いられる。

【0035】このように構成された文書記処理装置において、検索用情報の作成記憶、および文書データの記憶、検索、表示の動作を詳細に説明する。

【0036】図2は検索用情報の作成、記憶を行う処理の流れを示すものである。見出し部分は、文書データがODAなどの構造化文書であれば、構造から容易に検出できるが、文書データが構造化文書でなくとも内容を解析することで検出可能である。そこで文書データが図3(a)のような構造化文書である場合と図3(b)のような構造化されていない文書である場合のそれぞれについて説明する。

【0037】文書データは、まず、文書作成装置19により作成される。作成された文書データは文書データ記憶部11により記憶される。文書データを記憶するには、フロッピーディスク、ハードディスクなどの外部記憶装置を用いるが、データベース管理システムなど高度なシステムを用いることも可能である。文書データが文書データ記憶部11に記憶されると同時に、この文書データから見出し検出部121により見出部分の検出処理が行われるとともに、見出し複写部122による見出し

の内容の複写が行われる。

【0038】図2においてpは文書内容を示すポインタである。ポインタの指し示す対象は、構造化文書の場合、図3(a)に示されるように木構造のノードである。

【0039】まずステップS21によりpが文書の開始点を指し示すように初期化される。開始点は図3(a)の木構造のルートに当たる となる。

【0040】続いて、ステップS22により現在pが指し示している部分が見出しであるか否かを調べる。ここで、関数typeはpの指し示している対象の種類を調べ、それが見出しであれば見出しであることを表す値を返す。構造化文書では木構造の各ノードに、そのノードの種類、例えば表題、見出し、図形、本文などを区別する値が設定されているので、関数typeはこの値を調べることによってpが見出しであるか否かを知ることができる。図3(a)では、種類が見出しであるノードを二重丸で示している。

【0041】ステップS22によりpが見出しであると判定された場合、ステップS23において見出し複写部12によりその内容が複写される。複写された見出しは文書データを指し示す情報との対応表の形式にして検索用情報として検索用見出し記憶部13に記憶される。図3(a)に示すような構造化文書ではpが見出しである場合にはその子は内容部であり、見出しの内容となる文字列がこの内容部に格納されている。よって、複写対象はこの内容部に格納されている文字列となる。

【0042】ステップS22によりpが見出しでないと判定された場合、または、ステップS23により内容の複写が終了した後に、ステップS24により $p := next(p)$ によりpが次のノードを指し示すようにする。ここで、次のノードとは、pに子がある場合にはその最も左の子となるノードであり、子がない場合で右隣の兄弟があればそのノード、右隣の兄弟がなければ親の右隣の兄弟となるノードである。親の右隣の兄弟がない場合にはさらに親の右隣の兄弟を調べ、最初に見つかったノードとする。どの親にも右隣の兄弟がない場合には文書構造をすべて検査し終わったことになるのでpに終了したことを表すnilを設定する。

【0043】続いて、pの値を調べ、nilであれば終了し、nilでなければステップS22に戻って繰り返して処理を続ける。

【0044】次に、図3(b)のような構造化されていない文書に対する処理について説明する。ポインタpの指し示す対象は、構造化されていない文書では図3(b)に示されるように文字または行の先頭位置である。

【0045】まず、図2の処理フローのステップS21によりpが文書の開始点を指し示すように初期化される。開始点は文書の最初の行、文字である。

【0046】続いてステップS22により現在pが指し示している部分が見出しであるか否かを調べる。関数typeは、構造化されていない文書では、pの指し示している部分が、例えば数字で始まる、または、前後の行と異なるフォントが用いられているなどの条件を満たすか否かで、対象となっている部分が見出しであるか否かを調べる。

【0047】ステップS22によりpが見出しであると判定された場合、ステップS23によりその内容が複写される。複写された見出しは文書データを指し示す情報との対応表の形式で検索用情報として見出し記憶部13に記憶される。複写の範囲もステップS22において関数typeが用いたものと同様の条件を用いて決定する。

【0048】ステップS22によりpが見出しでないと判定された場合、または、ステップS23により内容の複写が終了した後にステップS24により $p := \text{next}(p)$ によりpが次の行、文字を指し示すようにする。次の行、文字がない場合には文書すべて検査し終わったことになるのでpに終了したことを表すnilを設定する。

【0049】続いてステップS25でpの値を調べ、nilであれば終了し、nilでなければステップS22に戻って繰り返し処理を続ける。

【0050】図4は、見出しと文書の対応関係を表す対応表の実現方法を示している。ここで、41は見出しと文書の対応を表す対応表であり、表の左側の列に見出し複写部122により複写された見出しを表す文字列43が格納され、右側の列に対応する文書データを表す情報42が格納される。44～47は文書データ記憶部11により記憶された文書データである。対応表41と文書データ44～47は同一のファイルに記憶してもよいが、別々のファイルに記憶しさらに文書データ44～47を個別のファイルに記憶することもできる。

【0051】対応表41と文書データ44～47が同一のファイルに記憶される場合、対応表41の右側の列には対応する文書データのファイル内での場所をポインタとして記憶することができる。

【0052】また、対応表41と文書データ44～47が別々のファイルに記憶されている場合には、対応表41の右側の列には対応する文書データが記憶されているファイルの名前とそのファイル中の場所を記憶する。

【0053】さらに文書データ44～47が個別のファイルに記憶される場合には、対応表41の右側の列には対応する文書データが記憶されているファイルの名前を記録する。

【0054】また、図4では同じ見出しに対して、対応表41に個別の行を作成しているが、同じ見出しを一行にまとめ、複数の文書データを対応付けることも可能である。

【0055】このようにして記憶された文書データから特定の文書データを、前記作成・記憶された検索用見出し情報により、取り出す検索処理について説明する。図5はその検索処理の流れを示す図である。

【0056】検索する文字列をユーザが入力するために、文書検索ユーザインタフェース15を用いる。入力する文字列は単一の完成された文字列でもよいが、正規表現などにより複数の文字列を表すものでもよい。また、検索する文字列を複数指定しそれらをAND・ORで統合することで、指定された複数の文字列を同時に含んでいる文書や一つでも含んでいる文書などの検索を指示することも可能である。

【0057】図5は検索処理のフローを示すものである。図5においてcは問い合わせの条件を表すデータ、iは対応表41の行を示す整数型の変数、uは対応表41から指されている文書で条件を満たすものを記憶する集合型の変数である。

【0058】ステップS51でまず問い合わせの条件を表すデータをcに代入する。条件を表すデータは、例えば正規表現により表すことができる。ここでは説明のため、正規表現「 $\cdot *aa \cdot * | \cdot *x$ 」で『「aa」を含むまたは「x」で終わる文字列』を指定する。

【0059】ステップS52でi、uを初期化する。次にステップS53で対応表のi行目である「対応表[i]」に記憶されている文字列が条件cを満たすか否かを調べる。iが1のときには対応表41の1行目を調べる。

【0060】ステップS53により調べた結果、条件を満たす場合にはステップS54により「対応表[i]」から指されている文書をuに追加する。図4の例では、対応表の1行目に記憶されている文字列「aa」が条件cの『「aa」を含む』を満たすので、文書データを指し示す情報をuに追加する。

【0061】ステップS53の判定の結果、条件を満たさないとされた場合、またはステップS54が終了した後に、ステップS55によりiの値が1だけ増える。

【0062】続いてステップS56によりその値が対応表41の大きさを越えていないか調べられ、越えていない場合にはステップS53にもどり、iが対応表41の大きさを越えるまで繰り返される。

【0063】図4の例では、この処理が終了した後に、uには文書データ44、45、47を指し示す情報が記憶される。これらの文書の一覧が文書検索ユーザインタフェース15により表示され、ユーザが文書指定部16によりその一覧の中から目的の文書の一つまたは複数指定すると、その指定された文書が文書取り出し部17により文書データ記憶部11から取り出され、文書表示装置18により表示される。

【0064】なお、以上の実施例では、各見出しを検出

するごとに、対応する見出しの内容を複写し、検索用情報として登録する例を説明したが、一つの文書について見出しを検出した見出し表に登録してゆき、その見出し表への登録がすべて終わった後に、その見出し表に基づいて見出しと文書を指し示す情報との対応表を作成するように変更実施することもできる。図6は見出し表と文書データとの関係を示す図であり、見出し表には検出した見出しの位置を示すポイントを順次格納する。

【0065】以上説明したように、本実施例は文書の見出しとなる文字列を検索用の情報として自動的に抽出し、登録するようにしたので、文書データを保存する際に、従来技術のように別途キーワードなどを指示する煩わしいキーワード抽出作業を必要とせず、文書データを容易に保存することができる。かつ検索時には目的とする文書を迅速かつ容易に取り出すことができる。

【0066】(第2の実施例)第1の実施例は、検索されたデータを文書単位で取り出して表示するものであるが、第2の実施例は検索されたデータを文書部品単位で取り出して文書作成に利用するようにしたものである。

【0067】図7は本発明の第2の実施例の文書処理装置の構成を示すブロック図である。この装置は、図7に示すように、文書部品記憶部71、検索用情報作成部72、検索用見出し記憶部73、見出し検索部74、文書部品検索ユーザインタフェース75、文書部品指定部76、文書部品取り出し部77、文書作成装置78および文書部品作成装置79からなっている。

【0068】文書部品記憶部71は文書部品作成装置79により作成された文書部品を記憶するものである。

【0069】検索用情報作成部72は、文書部品作成装置79からの文書部品を文書部品記憶部71に記憶する際に、文書部品に含まれる見出しから検索用情報を作成するものであり、文書部品作成装置79により作成された文書部品を解析し、章、節、図表の見出しとなる部分を検出する見出し検出部721と、検出した見出しの文字列を複写し文書部品との対応を付けて出力する見出し複写部722とを有している。

【0070】検索用見出し記憶部73は、検索用情報作成部72により作成された、見出しとなる文字列と文書部品記憶部71に記憶される文書部品との対応付けを表す検索用情報を記憶するものである。

【0071】見出し検索部74は文書部品を検索する際に用いられ、文書部品検索ユーザインタフェース75により与えられた文字列と同じまたはそれを含む見出しを検索用見出し記憶部73から検索し、該当する見出しを含む文書部品の一覧を作成するものである。

【0072】文書部品検索ユーザインタフェース75は、見出し検索部74に渡す文字列をユーザが入力したり、与えられた文書部品の一覧を表示しユーザから目的とする文書部品を文書部品指定部76により指定させるための入出力を司るものである。

【0073】文書部品指定部76は見出し検索部74より作成された文書部品の一覧を文書部品検索ユーザインタフェース75により表示し、取り出すべき文書部品をユーザに選択させるものである。

【0074】文書部品取り出し部74は、文書部品指定部76により指定された文書部品を文書部品記憶部71から取り出すものである。

【0075】文書作成装置78は文書部品取り出し部77により取り出された文書部品を利用して文書データを作成するものである。

【0076】文書部品作成装置79は文書部品を作成するものである。

【0077】以上のように構成された第2の実施例の文書記処理装置において、検索用情報の作成記憶の動作を説明する。文書部品は、まず文書部品作成装置79により作成される。作成された文書部品は文書部品記憶71により記憶される。文書部品を記憶するには、フロッピーディスク、ハードディスクなどの外部記憶装置を用いるが、データベース管理システムなど高度なシステムを用いることも可能である。文書部品が文書部品記憶部71に記憶されると同時にこの文書部品から見出し検出部721により見出し部分が検出される。見出し部分は、文書部品がODAなどの構造化文書に用いることができる形式で作成されたものであれば、その構造から容易に検出することができる。文書部品が構造化文書用のものでなくとも内容を解析することにより検出可能である。例えば文書部品がテキストであれば数字で始まる、または、前後の行と異なるフォントが用いられているなど、文書部品が図や表であれば「図」「表」の文字を含む最初または最後の行であるというような条件を満たすか否かで、対象となっている部分が見出しであるか否かを調べる。

【0078】検出した見出しは、文書部品中の場所を指すポイントなどによって示され、表形式で一時記憶される。図8は、検出された見出しの記憶方法を示している。ここで81は見出し表であり検出された見出しを示すポイントが記憶される。82、83、84は文書部品であり、文書部品作成装置79により作成されたものである。検出された見出しは図8のようにポイントなどによって図中の文書部品82～84中の場所で指し示し、図中81の見出し表に登録しておく。

【0079】このようにして得られた見出し表81と文書部品82、83、84…から、見出しとなる文字列が見出し複写部722により複写され、見出しと文書部品記憶部71により記憶された文書部品の対応を表す対応表として見出し記憶部81により記憶される。

【0080】図9は、見出しと文書部品の対応を表す対応表の実現方法を示している。ここで、91は見出しと文書部品の対応を表す対応表であり、表の左側の列に見出し複写部722により複写された見出しを表す文字列



93が格納され、右側の列に対応する文書部品を示す情報92が格納される。94～99は文書部品記憶部71により記憶された文書部品である。対応表91と文書部品94～99は同一のファイルに記憶してもよいが、別々のファイルに記憶しさらに文書部品94～99を個別のファイルに記憶することもできる。対応表91と文書部品94～99が同一のファイルに記憶される場合、対応表91の右側の列には対応する文書部品のファイル内での場所をポインタとして記憶することができる。

【0081】また、対応表91と文書部品94～99が別々のファイルに記憶されている場合には、対応表91の右側の列には対応する文書部品が記憶されているファイルの名前とそのファイルのなかの場所を記憶する。さらに文書部品94～99が個別のファイルに記憶される場合には、対応表91の右側の列には対応する文書部品が記憶されているファイルの名前を記録する。

【0082】なお、図9では同じ見出しに対して、対応表91に個別の行を作成しているが、同じ見出しを一行にまとめ、複数の文書部品を対応付けることも可能である。また、上記の実施例では見出し表を作成し、その後その見出し表を用いて対応表を作成する場合を説明したが、見出し表を作成せずに第1の実施例の図2で示したように見出しを検出する毎に対応表に直接に登録するようにしてもよい。

【0083】このようにして作成・記憶された検索用見出し情報即ち対応表により、文書部品記憶部74に記憶された文書部品群から所望の文書部品を取り出す検索処理は、図5のフローチャートにより説明した第1の実施例における検索処理とほぼ同じである。対応表の探索により、第1の実施例では文書の一覧を得るのに対し、第2の実施例では文書部品の一覧を得る点が相違するのみである。

【0084】得られた文書部品の一覧は文書部品検索ユーザインタフェース75により表示し、ユーザが目的とする文書部品を文書部品指定部76により指定できるようにする。ここで文書検索ユーザインタフェース75は単一の文書部品を指定させるのもよいが、複数の文書部品を指定できるようにすることも可能である。ユーザが文書部品検索ユーザインタフェース75と文書部品指定部76により一つまたは複数の文書部品が指定されると、指定された文書部品は文書部品取り出し部77により文書部品記憶部71から取り出され、文書作成装置78に取り込まれる。

【0085】以上説明したように、本実施例は文書部品の見出しとなる文字列を文書部品と対応させて検索用情報として検索用見出し記憶部73に自動的に登録するようにしたので、文書部品を保存する際に、別途キーワードなどを人手により指示する必要がなく、大量の文書部品であっても容易に保存し、かつ検索時には目的とする文書部品を迅速かつ容易に取り出すことができる。

【0086】(第3の実施例)図10は本発明の第3の実施例の文書処理装置の構成を示すブロック図である。この文書処理装置は、図10に示すように、文書データ記憶部101、検索用情報作成部102、検索用見出し記憶部103、見出し検索部104、文書検索ユーザインタフェース105、文書指定部106、文書取り出し部107、文書データ編集部108および見出し編集記録部109からなっている。

【0087】この文書処理装置は文書データ編集部108により文書データ記憶部101に格納されている文書データの編集を行うようにしたものであり、編集により文書データが更新された時に、その更新された文書データに検索用見出し記憶部103の内容を整合させるようにしたものである。第1の実施例とは、文書データ記憶部101から検索された文書データを編集する文書データ編集部108と、文書データ編集部108により追加、変更、削除された見出しを抽出する見出し編集記録部109と、見出し編集記録部109により記録された見出しの追加、変更、削除を検索用見出し記憶部103に反映させる見出し更新部1023とを設けた点において相違し、その他の構成要素は第1の実施例と同じである。従って、文書データおよび検索用見出し情報の新規な登録処理、および文書データの検索処理については説明を省略し、前記相違点に関する部分即ち検索用見出し情報の更新処理に重点を置いてその動作を説明する。

【0088】文書データ編集部108による編集では見出しの追加、変更、削除が可能である。文書データ編集部108による見出しの追加、変更、削除は、見出し編集記録部109により検出され、追加、変更、削除の別に記録される。文書データ編集部108による編集が終了し、再び、文書データ記憶部101により記憶される際に、見出し更新部1023は、見出し編集記録部109により記録された見出しの追加、変更、削除の情報により検索用見出し記憶部103に格納されている対応表の更新を行なう。この更新処理において、対応表に、追加された見出しに対する行が追加され、変更された見出しに対する行の内容が変更され、削除された見出しに該当する行が削除される。以下に、その詳細な処理を説明する。

【0089】図11は見出し検出部1021により検出された見出しの記憶方法を示している。図中114は編集対象として指定された文書データを文書データ記憶部101から文書取り出し部107により取り出して編集用作業領域に置かれた編集対象文書データである。111は追加見出し表であり、文書データ編集部108による編集で追加された見出しを指し示すポインタが記憶される。112は変更見出し表であり、文書データ114に対し文書データ編集部108による編集で変更された見出しを指し示すポインタが記憶される。113は削除見出し表であり、編集対象文書データ114から文書編

集部108による編集で削除された見出しを指し示すポイントが記憶される。この図では、第1章の見出しの内容が「x x x」から「x x x y y y」に変更され、第2章の「y y y」が削除され、最後に「p p p」と「m m m」が追加された状態を示している。

【0090】見出し編集記録部109は、追加見出し表111、変更見出し表112、削除見出し表113を図12のフローチャートに示された、以下の手順で作成する。

【0091】見出し編集記録部109は、まず、文書データ編集部108による編集操作の種類を調べる（ステップS121、S123、S126）。

【0092】文書データ編集部108による編集操作が見出しの追加であれば、その見出しへのポイントを追加見出し表111に記憶する（ステップS122）。

【0093】文書データ編集部108による編集操作が見出しの変更であれば、その見出しへのポイントが追加見出し表111に記憶されているか調べ（ステップS124）、追加見出し表111に記載されていない場合のみ変更見出し表112に記憶する（ステップS125）。

【0094】文書データ編集部108による編集操作が見出しの削除であれば、まず、その見出しへのポイントが追加見出し表111に記憶されているか調べ（ステップS127）、追加見出し表111に記憶されている場合は追加見出し表111から削除し（ステップS128）、そうでない場合のみ削除見出し表113に記憶し（ステップS129）、さらにその見出しへのポイントが変更見出し表112に記憶されているか調べ、（ステップS12A）、記憶されている場合は変更見出し表112からそのポイントを削除する（ステップS12B）。

【0095】文書編集部108により編集が終了すると、編集対象文書データ114が文書記憶部101に格納されている編集前のもとの文書データを更新する形で記憶される。その際、以上の手順により作成された追加見出し表111、変更見出し表112、削除見出し表113を用いて、見出し更新部1023は、見出し記憶部103に記憶されている対応表131を更新する。

【0096】図13は対応表の更新を説明するための図であり、対応表131と、文書データ記憶部101の更新後の文書データ132と、文書データ編集部108において編集操作された編集対象の文書データ133と、変更見出し表112および削除見出し表113に登録された文書データ133上の変更または削除の施された見出しの位置のポイント（右側）と更新後の文書データ132の対応す

る位置のポイント（左側）とを対応させた追加見出し対応表135との相互の関係が示されている。

【0097】この図13に示すように、編集見出し対応表134および追加見出し対応表135には、文書データ記憶部101の文書データ132が更新された時に、編集によって変更、削除または追加された見出しの文書データ133における位置と更新により文書データ132中に変更、削除または追加された見出しの位置との対応が登録される。

10 【0098】図13の状態から、図11の追加見出し表111、変更見出し表112、削除見出し表113を用いて、対応表131の内容を更新する。この更新は、削除、追加、変更の順に、それぞれ図14、図15、図16に示される手順で行われる。

【0099】図14においてiは削除見出し表の行を示す変数であり、ステップS141において1に初期化される。

20 【0100】続いてステップS142でiが削除見出し表113の大きさを越えていないか確認する。越えている場合には、削除見出し表113の全ての行を処理したことになるので、終了する。

【0101】越えていない場合には、ステップS143により削除見出し表113のi行目の内容を変数iに代入する。

30 【0102】続いてステップS144によりiにより指し示された削除された見出しの編集前の位置を示すポイントをpに代入する。削除された見出しの編集前の位置は図13の編集見出し対応表134から、右側がiと同じである行を探し、その行の左側の値を参照することにより得ることができる。

【0103】次にステップS145において、対応表131からpと同じ位置を示すポイントを記憶している行を削除する。最後に、ステップS146でiの値を1だけ増加させてステップS142に戻る。

【0104】図15においてiは追加見出し表111の行を指し示す変数であり、ステップS151により1に初期化される。

40 【0105】続いてステップS152でiが追加見出し表111の大きさを越えていないか確認する。越えている場合には、追加見出し表の全ての行を処理したことになるので、終了する。

【0106】越えていない場合には、ステップS153により追加見出し表111のi行目の内容を変数iに代入する。

50 【0107】続いてステップS154において、iにより指し示された追加された見出しの更新後の位置を示すポイントpに代入する。追加された見出しの更新後の位置は図13の追加見出し対応表135から、右側がiと同じである行を探し、その行の左側の値を参照することにより得ることができる。

【0108】次にステップS155で、対応表131に、見出しの内容とpと同じ位置を示すポイントを記憶する行を追加する。最後にステップS156でiの値を1だけ増加させてステップS152に戻る。

【0109】図16において、iは変更見出し表112の行を指し示す変数であり、ステップS161により1に初期化される。

【0110】続いてステップS162でiが変更見出し表112の大きさを越えていない確認する。越えている場合には、変更見出し表112の全ての行を処理したことになるので、終了する。

【0111】越えていない場合には、ステップS163により変更見出し表112のi行目の内容を変数iに代入する。

【0112】続いてステップS164において、iにより指し示された変更された見出しの編集前の位置を示すポイントをpに代入する。変更された見出しの更新後の位置は図13の編集見出し対応表134から、右側がiと同じである行を探し、その行の左側の値を参照することで得ることができる。

【0113】次にステップS165で、対応表131に、見出しの内容とpと同じ位置を示すポイントを記憶する行の左側の内容、変更された見出しの内容で更新する。

【0114】最後にステップS166でiの値を1だけ増加させてステップS162に戻る。

【0115】本実施例によれば、文書データ編集部108により作成され、文書データ記憶部101により記憶しようとする文書の内容を表すデータから、見出し検出部1021により見出しを検出し、その見出しと文書データ記憶部101に記憶された文書の内容を表すデータとの対応を見出し複写部1022により作成し、検索用見出し記憶部103に記憶させるようにしたので、文書を登録する際に、従来のように検索用のキーワードを付加する作業を必要とせず、検索用の情報を自動的に生成することができる。また、検索用見出し記憶部103が特定の記憶領域に設けられた場合、参照すべき検索用の情報の領域を局所化することが可能となるので、検索速度を高速化することができる。

【0116】また、本実施例において検索時には、与えられた文字列と同じかまたはその文字列を含む文字列を検索用見出し記憶部103に格納した検索用情報の対応表131により検索し、その文字列に対応付けられている文書の一覧を提示し、その一覧から目的とする文書を選択することにより所望の文書を取り出すことができる。

【0117】さらに本実施例によれば、文書データ記憶部101により記憶された文書の内容を表すデータを文書データ編集部108により編集する際に、見出し編集記録部109により追加、更新、削除された見出しを記

録しておき、再度、文書データ記憶部101により文書データを記憶する際に、見出し更新部1023により、見出しの編集記録部109の記録にしたがって検索用見出し記憶部103の内容を更新することができる。

【0118】また、検索用見出し記憶部103の内容の更新の際に、削除された見出しの更新を最初に行うことで、削除された見出しのための領域が新たに追加された別の見出しのための領域として再利用された場合にも正確に更新を行うことができる。

10 【00119】(第4の実施例) 前述の第2の実施例は、文書部品を格納し、その文書部品を高速に検索するための検索用情報を文書部品の見出しから抽出するようにしたものである。その第2の実施例は文書部品を単位として登録し、文書部品単位で再利用するのに対し、第4の実施例は、文書全体を登録し、その文書を構成する文書部品を文書部品単位で検索し再利用の対象とするものである。

20 【0120】図17は本発明の第4の実施例の文書処理装置の構成を示すブロック図である。この装置は、図17に示すように、構造化文書記憶部171、検索用情報作成部172、部品検索用情報記憶部173、文書部品検索部174、文書部品検索ユーザインタフェース175、文書部品指定部176、文書部品取り出し部177、および文書作成装置178からなっている。

【0121】構造化文書記憶部171は文書作成装置178により作成された構造化文書を記憶するものである。

30 【0122】検索用情報作成部172は、文書作成装置178からの構造化文書を構造化文書記憶部171に記憶する際に、文書部品を単位として検索するための検索用情報を作成するものであり、文書作成装置178により作成された構造化文書の文書構造を解析し、章、節、図表の見出しとなる部分を検出する見出し検出部1721と、検出した見出しの文字列と文書部品との対応を示す部品検索用オブジェクトを生成する部品検索用情報作成部1722とを有している。

【0123】部品検索用情報記憶部173は、検索用情報作成部172により作成された部品検索用情報を記憶するものである。

40 【0124】文書部品検索部174は文書部品を検索する際に用いられ、文書部品検索ユーザインタフェース175により与えられた文字列と同じまたはそれを含む見出しを部品検索用情報記憶部173から検索し、該当する見出しを含む文書部品の一覧を作成するものである。

【0125】文書部品検索ユーザインタフェース175は、文書部品検索部174に渡す文字列をユーザが入力したり、与えられた文書部品の一覧を表示しユーザから目的とする文書部品を文書部品指定部176により指定させるための入出力を司るものである。

50 【0126】文書部品指定部176は文書部品検索部1

74より作成された文書部品の一覧を文書部品検索ユーザインタフェース175により表示し、取り出すべき文書部品をユーザに選択させるものである。

【0127】文書部品取り出し部174は、文書部品指定部176により指定された文書部品を構造化文書記憶部171から取り出すものである。

【0128】文書作成装置178は文書部品取り出し部177により取り出された文書部品を利用して文書データを作成するものである。

【0129】本実施例の装置で取り扱う構造化文書の具体例を図18に示し、その文書の論理構造を図19に示す。この構造化文書は3つの章を含み、「1章」はそのタイトルの内容と2つの節、すなわち「1. 1節」および「1. 2節」からなっている。「1. 1節」は、そのタイトルの内容と本文の内容と「図1」からからなっている。「図1」はその図形のキャプションと内容である図形情報からなっている。

【0130】構造化文書記憶部171では、文書の論理構造を上述した図19に示すような構造で保持する。部品検索用情報記憶部173は図20に示すような文書部品の単位ごとに検索対象となる見出し（タイトルあるいはキャプション）と文書部品が組みとなった検索用のオブジェクトを要素とする集合を部品検索用情報として保持する。

【0131】図21は、図18に示された文書が構造化文書記憶部171へ格納される際に、検索用情報作成部172により生成される検索用オブジェクトの具体例を示すものである。検索用オブジェクトの値として各文書部品に対応する見出しと文書部品のポイントの対が生成される。

【0132】第1の実施例の説明に用いた図2は、本実施例の検索用情報の作成、記憶を行う処理の流れの説明にも用いることができる。図2においてpは文書内容を示すポイントである。ポイントの指し示す対象は、構造化文書の本構造のノードである。

【0133】まずステップS21によりpが文書の開始点を指し示すように初期化される。開始点は図3(a)の本構造のルートに当たる となる。

【0134】続いて、ステップS22により現在pが指し示している部分が見出しであるか否かを調べる。ここで、関数typeはpの指し示している対象の種類を調べ、それが見出しであれば見出しであることを表す値を返す。構造化文書では本構造の各ノードに、そのノードの種類、例えば表題、見出し、図形、本文などを区別する値が設定されているので、関数typeはこの値を調べることによってpが見出しであるか否かを知ることができる。図3(a)では、種類が見出しであるノードを二重丸で示している。

【0135】ステップS22によりpが見出しであると判定された場合、ステップS23において部品検索用情

報作成部1722によりそのpの内容とその位置を指し示す情報との対応表の形式にした部品検索用情報が作成され、部品検索用情報記憶部173に記憶される。図3(a)に示すような構造化文書ではpが見出しである場合にはその子は内容部であり、見出しの内容となる文字列がこの内容部に格納されている。

【0136】ステップS22によりpが見出しでないと判定された場合、または、ステップS23により部品検索用情報の作成、記憶が終了した後に、ステップS24により $p := next(p)$ によりpが次のノードを指し示すようにする。

【0137】続いて、pの値を調べ、nilであれば終了し、nilでなければステップS22に戻って繰り返して処理を続ける。

【0138】このようにして記憶された作成・記憶された部品検索用情報を用いて、構造化文書記憶部171から所望の文書部品を取り出す検索処理について説明する。第1の実施例の説明に用いた検索処理の流れを示す図5は本第4の実施例にも適用できる。

【0139】図5においてcは問い合わせの条件を表すデータ、iは対応表41の行を示す整数型の変数、uは対応表41から指されている文書部品で条件を満たすものを記憶する集合型の変数である。

【0140】ステップS51でまず問い合わせの条件を表すデータをcに代入する。

【0141】ステップS52でi、uを初期化する。次にステップS53で対応表のi行目である「対応表[i]」に記憶されている文字列が条件cを満たすか否かを調べる。iが1のときには対応表41の1行目を調べる。

【0142】ステップS53により調べた結果、条件を満たす場合にはステップS54により「対応表[i]」から指されている文書部品をuに追加する。

【0143】ステップS53の判定の結果、条件を満たさないとされた場合、またはステップS54が終了した後に、ステップS55によりiの値が1だけ増える。

【0144】続いてステップS56によりその値が対応表41の大きさを越えていないか調べられ、越えていない場合にはステップS53にもどり、iが対応表41の大きさを越えるまで繰り返される。

【0145】この処理が終了した後に、uには条件を満たす文書部品を指し示す情報が記憶されている。これらの文書部品の一覧が文書部品検索ユーザインタフェース175により表示され、ユーザが文書部品指定部176によりその一覧の中から目的の文書部品を一つまたは複数指定すると、その指定された文書部品が文書部品取り出し部177により構造化文書記憶部11から取り出され、文書作成装置178により利用される。

【0146】次に、文書部品検索の具体例として図18、図19に示した文書が格納されている構造化文書記

億部171に対する検索の例を挙げる。

検索例1:「データベース」を含んでいる見出しを持つ部品を検索

検索結果:1章全部のサブツリー を含んでいる要素が検索結果の集合に入れられる。

検索例2:「QODB」を含んでいる見出しを持つ部品を検索

検索結果:1、2節のサブツリー を含んでいる要素が検索結果の集合に入れられる。

検索例3:「図1」を含んでいる見出しを持つ部品を検索

検索結果:図1のサブツリー を含んでいる要素が検索結果の集合に入れられる。

【0147】本実施例によれば、文書部品を選択的に格納するのではなく、文書そのものを格納することにより、既存文書を構成する全ての文書部品を再利用の対象とすることができる。さらに文書として格納しているので、格納時に、いちいち文書部品の指定をする手間がはぶける。また、文書全体を格納していることにより、従来技術（例えば、前記特許特開平2-148250号公報、特開平3-8087号公報）より領域を必要とするように見えるがそもそも保存する必要がある文書を格納しているので無駄な領域を使用している訳ではない。従来技術では、再利用のために登録した文書部品を含んでいる文書を保存している場合、文書とは別に再利用ための部品を2重に記憶しているので、本実施例の方がトータルでの必要とする領域も少なくてすむ。さらに、文書を単位として登録格納するので、文書部品を単位として登録格納する場合のように文書部品を別途作成する必要がない。また、本実施例によれば文書部品をそれが所属している文書を気にせずに文書部品として検索することが可能となる。なお、本実施例の変形例として、構造化文書記憶部171に記憶された文書に編集処理を施して元の文書を書き換えることができるように構成した場合には、図10に示す第3の実施例と同様に検索用見出し記憶部の内容を更新するための手段が必要となる。即ち、この変形例を実現するためには図10の文書データ編集部108、見出し編集記録部109および見出し更新部1023にそれぞれ対応する構成を図17の構成に付加すればよい。

【00148】（第5の実施例）この実施例は特定の文書部品を各構造化文書から取り出して、一つの文書に組み合せて提示することのできる文書処理装置に関する。図22はこの実施例の機能ブロック図である。

【0149】この処理装置は、図22に示すように、構造化文書記憶部221と、検索条件入力部222と、文書構造検索部223と、仮想文書作成部224と、文書表示部225を備えている。

【0150】構造化文書記憶部221は文書を格納する記憶部であり、その文書としてはODA形式の構造化文

書を取り扱うものとする。構造化文書はいくつかの文書部品から構成されている。本実施例で扱う文書部品の単位は、章、節、図形、表、イメージ等を扱うものとする。図23は構造化文書の例を示し、図24はその構造木を示すものである。

【0151】検索条件入力部222は、取り出すべき特定の文書部品の条件を入力するものである。

【0152】文書構造検索部223は、構造化文書の形で格納された各文書から検索条件入力部222により入力された条件を満たす特定の文書部品を構造化文書記憶部221から検索するものであり、文書ルート取得部2231、文書部品選択部2232を有している。文書ルート取得部2231は構造化文書記憶部221に格納されている文書の構造木のルートを順次取り出す処理を行うものである。文書部品選択部2232は文書ルート取得部2231で取り出した文書ルートから構造木を辿って、検索条件入力部222で取得した文書部品検索条件を満たす文書部品を選択する処理を行うものである。

【0153】仮想文書作成部224は、文書構造検索部223で特定された文書部品を利用して文書を作成するものである。

【0154】図25は文書部品を検索する処理のフローを示すものである。

(1) 最初に検索条件入力部11により特定する部品の検索条件を入力する（ステップS251）。ここで文書部品検索条件は、各文書中の特定の文書部品を抽出するための条件である。例えば、文書部品検索条件には、文書部品の見出しを指定することによって特定するための例として見出し文字列が「Abstract」であるという条件がある。また、定型フォームを持つ文書において決まった位置に出現する文書部品を指定するためにその位置を条件としてもよい。このように文書の構造に依存する条件を用いることができる。

【0155】(2) 照合すべき対象の文書があるか否かを調べ（ステップS252）、照合すべき対象の文書があれば、文書ルート取得部2231によって構造化文書記憶部221から文書を構成するツリーのルートの一つを取り出す（ステップS253）。すべて照合が済んで処理すべき対象の文書がない場合は(5)の処理へ進む。

【0156】(3) そのツリーに照合が済んでいない文書部品があるか否かを調べ（ステップS254）、あれば文書部品選択部122は(2)で得られたルートから文書部品のツリーを辿り、文書部品の一つを取り出す（ステップS255）。取り出すべき文書部品がそのツリーにない場合は(2)の処理に戻る。ここで定型フォームを持つ文書などを検索対象とする場合、ツリーにおけるある位置の文書部品のみを候補として取り出すといった手法もある。

【0157】(4) 各々の文書部品が文書部品検索条

件を満たすかどうかを判定する（ステップS256）。例えば、文書部品検索条件として文書部品の見出しに含まれるべき文字列が指定されているときは、（3）で取り出した文書部品に対応する見出しの文字列と文書部品検索条件の文字列とを比較し、一致した場合にはその文書部品が条件を満たす文書部品として仮想文書作成部224へ渡して（ステップS257）、（2）の処理へ戻る。文書部品検索条件を満たさない場合、（3）の処理へ戻る。

【0158】（5）仮想文書作成部224は検索の結果得られた文書部品群を内部メモリ領域中において組み合わせて一つの仮想的に一つの文書を生成する（ステップS258）。即ち、仮想文書作製部224は、図26に示すように、仮想文書作成部224の作業用の内部メモリに文書論理構造に関する文書オブジェクト261を生成し、文書構造検索部223により特定された個々の文書部品を構造化文書記憶部221から読み込み、生成した文書オブジェクトに繋ぎ合わせていき、文書部品の関係（親子関係）を構築する。

【0159】文書表示部225は、仮想文書作成部224が作成した文書情報を可視化する。図27に論文のAbstractを表示した仮想文書の一例を示す。この例の場合、論文の集まりからそのアブストラクトのみを抽出し、そのアブストラクトと文書名を組みにした段落で構成される仮想文書が作成され、表示されたものである。ユーザはこの仮想文書を読むことにより、所望の論文を探すことができる。

【0160】本実施例によれば、2次記憶装置である構造化文書記憶部221に格納された既存文書をベースにして、内部メモリ中に仮想文書を作成しているので2次記憶装置の領域を必要としない。さらに、もし仮想文書を保管したい場合には、内部メモリ中に記憶されている仮想文書情報を構造化文書記憶部221に格納することにより保管できる。さらに、文書の共有機能を利用すれば、各論文のアブストラクトとこの文書の各々の項目を共有することにより、記憶領域の節約、論文が編集されたときの内容の更新の追従を行い、つねに新しい情報を維持できる。また、本実施例は、文書の検索に用いるばかりではなく、既存文書のある部分を抽出して一つの別の文書として見ることを可能とするので、文書の構成要素を格納時の目的とは別の多目的に再利用することができるようになる。

【0161】なお、上記の例では文書部品の特定は、文書部品が持っている見出しがAbstractである文書部品を特定したが、文書部品の構造木における位置で特定してもよい。また、各文書ごとに、特定する文書部品を変えるようにしてもよい。

【0162】また、仮想文書を文書表示部225に表示するほかに、印刷部により印刷するようにすることもできる。なお、仮想文書をファイルとして構造化文書記憶

部221に保存することもできる。

【0163】

【発明の効果】本発明（請求項1）によれば、文書に含まれる見出しを見出し検出手段により抽出して、検索用情報を作成し、従来のキーワードの代わりに登録するようにしたので、キーワードを選定する煩雑な作業を要することなく容易に検索対象文書の登録が可能になる。かつ検索時には目的とする文書を迅速かつ容易に取り出すことができる。

10 【0164】本発明（請求項2）によれば、見出し編集記録手段に見出しの編集操作を記録し、その記録に基づいて検索用情報更新手段が検索用情報を更新するようにしたので、編集の結果が検索用情報に反映され、文書の検索を常に正確になるように維持できる。

【0165】本発明（請求項3）によれば、文書要素を選択的に格納するのではなく、文書そのものを格納し、検索は文書要素単位で取り出すようにしたことにより、既存文書を構成する全ての文書要素を再利用の対象とすることができる。さらに文書として格納しているので、20 格納時に、いちいち文書要素の指定をする手間が省くことができ、また、文書の編集のみを行えば文書要素を検索用に別途編集する必要がない。

【0166】本発明（請求項4）によれば、既存文書の所望の部分を抽出して一つの別の文書としてユーザに提示することができるので、ユーザは提示された文書を見て必要な文書の選択をすることができ、また、文書の要素を格納時の目的とは別の複数の目的に再利用することもできる。

【0167】本発明（請求項5）によれば、前記発明（請求項1）と同様に文書要素の登録が容易になると共に、30 文書全体を登録した文書情報からの文書要素の検索において、より高速の検索が可能となる。

【0168】本発明（請求項5）によれば、文書要素に含まれる見出しを見出し検出手段により抽出して、検索用情報を作成し、従来のキーワードの代わりに登録するようにしたので、キーワードを選定する煩雑な作業を要することなく容易に検索対象文書要素の登録が可能になる。

【図面の簡単な説明】

40 【図1】 第1の実施例の構成を示す図

【図2】 文書登録の処理フローを示す図

【図3】 （a）は構造化文書の例、（b）は非構造化文書の例

【図4】 対応表の例を示す図

【図5】 文書検索の処理フローを示す図

【図6】 見出し表の例を示す図

【図7】 第2の実施例の構成を示す図

【図8】 見出し表の例を示す図

【図9】 対応表の例を示す図

50 【図10】 第3の実施例の構成を示す図

【図11】 見出しの削除、変更、追加等の編集操作の記録を示す図

【図12】 編集操作の記録の処理フローを示す図

【図13】 更新箇所の対応表の例を示す図

【図14】 対応表（検索用情報）の見出しの削除の処理フローを示す図

【図15】 対応表の見出しの追加の処理フローを示す図

【図16】 対応表の見出しの変更の処理フローを示す図

【図17】 第4の実施例の構成を示す図

【図18】 構造化文書の例を示す図

【図19】 構造化文書の構造の例を示す図

【図20】 検索用オブジェクトの構造の例を示す図

【図21】 検索用オブジェクトの具体例を示す図

【図22】 第5の実施例の構成を示す図

【図23】 構造化文書の例を示す図

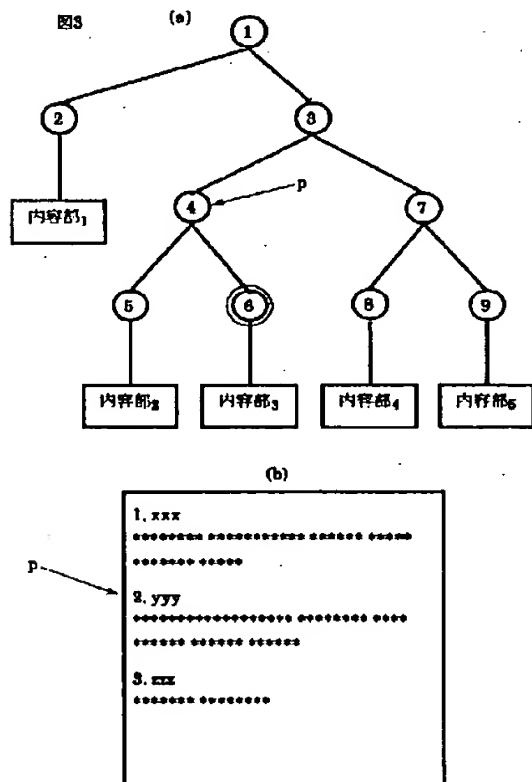
【図24】 構造化文書の構造の例を示す図

【図25】 第5の実施例の処理フローを示す図

【図26】 仮想文書と文書情報との関係を示す図

【図27】 仮想文書の表示例を示す図

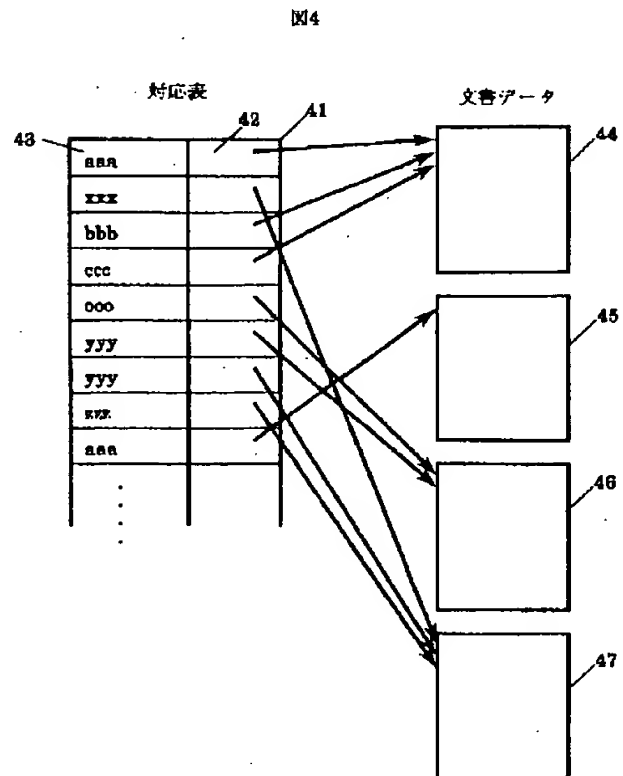
【図3】



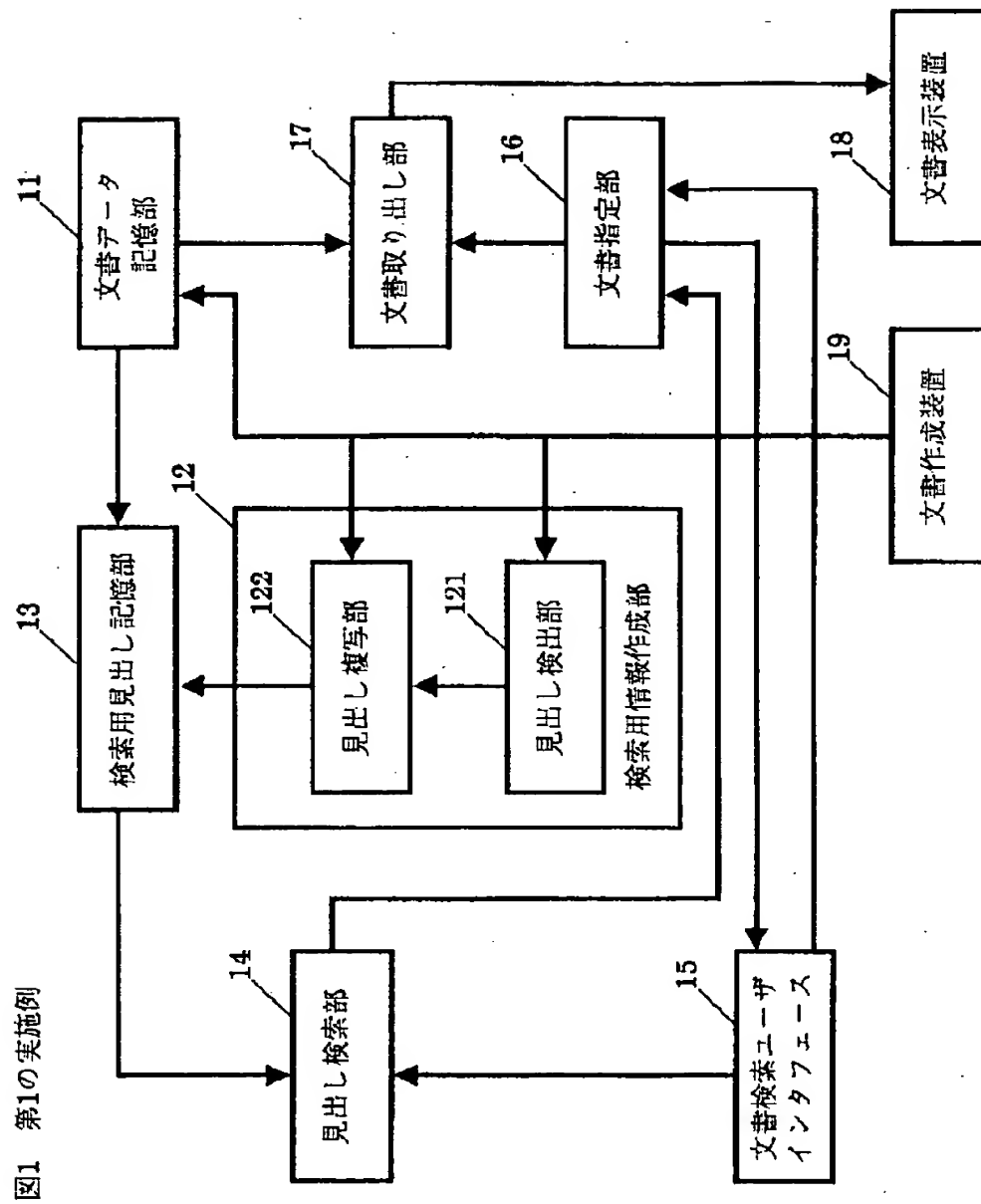
【符号の説明】

11、101…文書データ記憶部、12、72、102、172…検索用情報作成部、121、721、1021、1721…見出し検出部、122、722、1022…見出し複写部、1722…部品検索用情報作成部、13、73、103…検索用見出し記憶部、14、74、104…見出し検索部、15、105…文書検索ユーザインタフェース、16、106…文書指定部、17、107…文書取り出し部、18…文書表示装置、19…文書作成装置、71…文書部品記憶部、75、175…文書部品検索ユーザインタフェース、76…文書部品指定部、77…文書部品取り出し部、78…文書作成装置、79…文書部品作成装置、108…文書データ編集部、109…見出し編集記録部、1023…見出し更新部、171…構造化文書記憶部、173…部品検索用情報記憶部、174…文書部品検索部、221…構造化文書記憶部、222…検索条件入力部、223…文書構造検索部、2231…文書ルート取得部、2232…文書部品選択部、224…仮想文書作成部、225…文書表示部。

【図4】



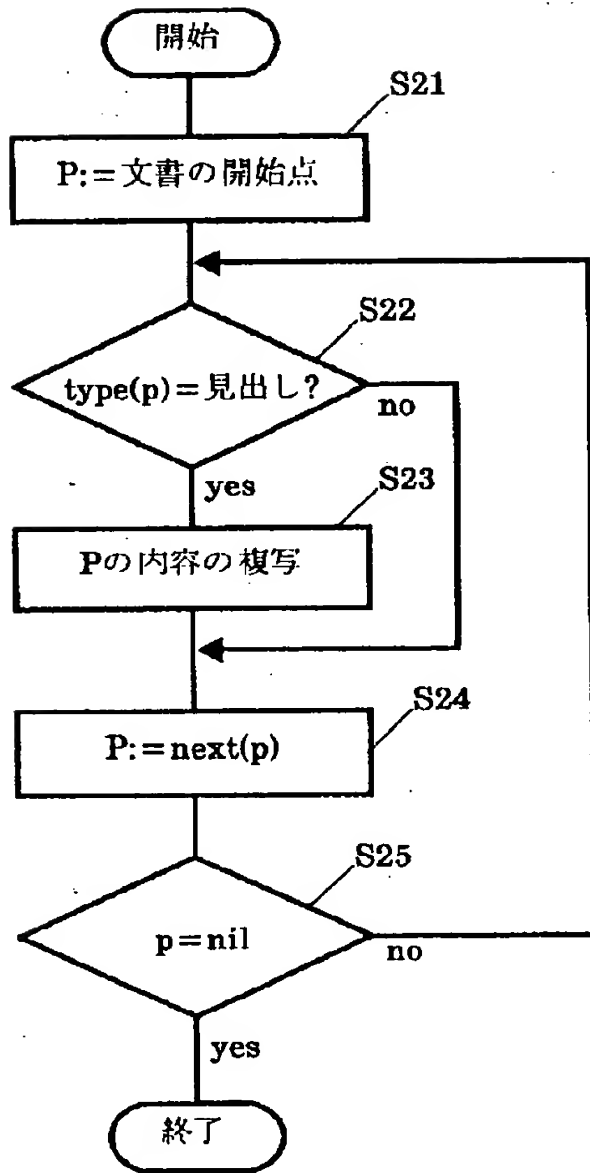
【図 1】





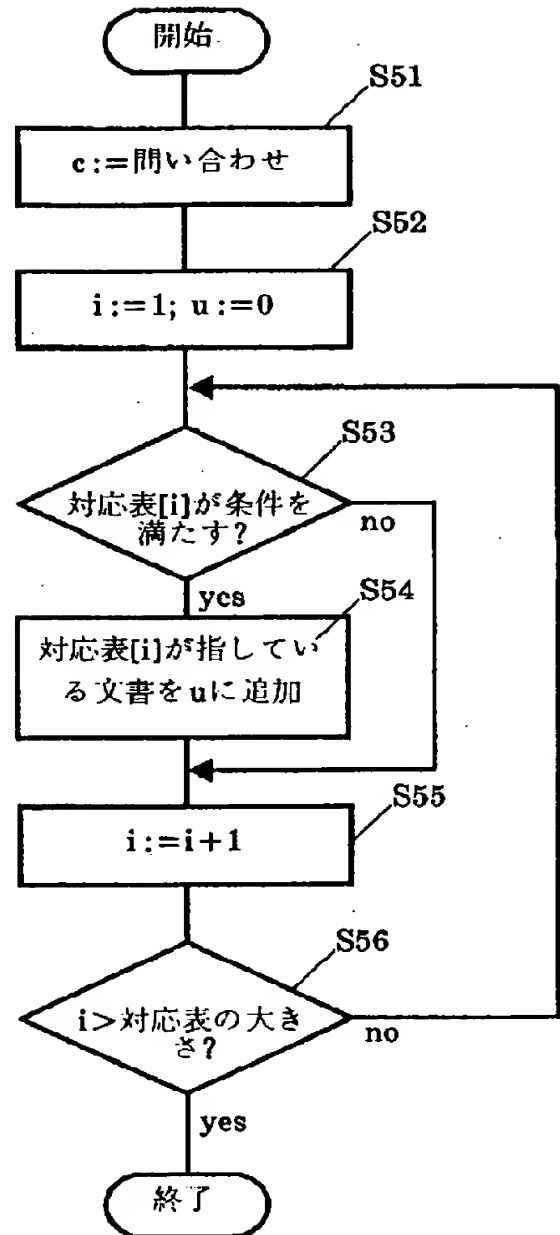
【図 2】

図2



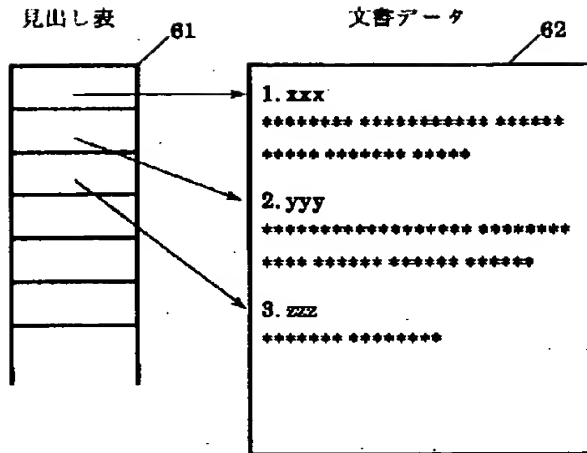
【図 5】

図5



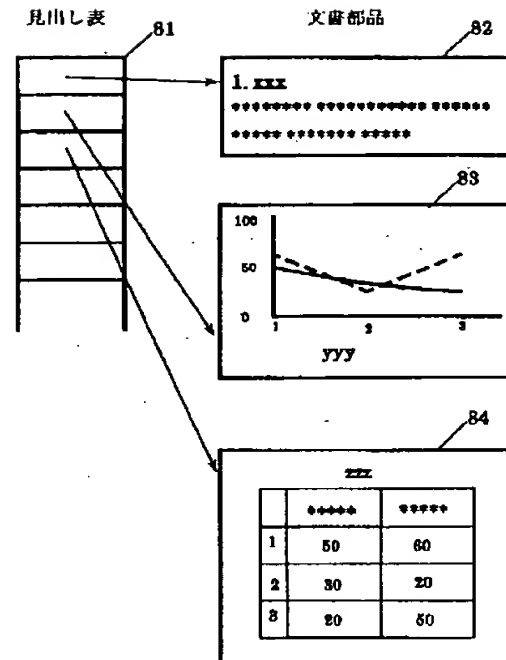
【図6】

図6



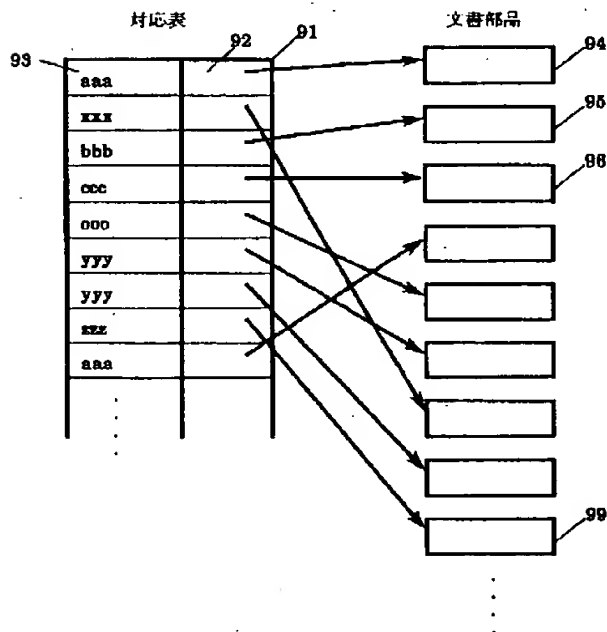
【図8】

図8



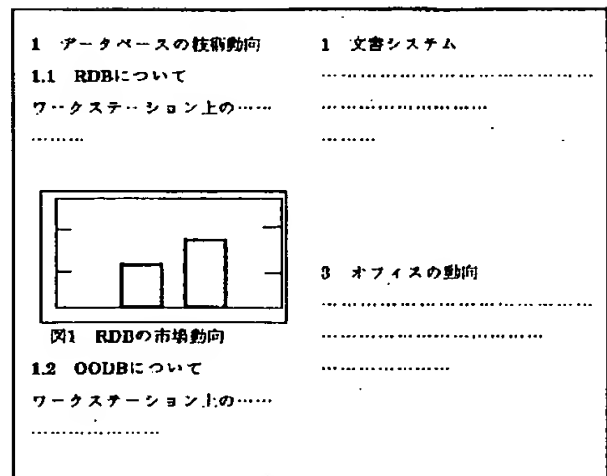
【図9】

図9

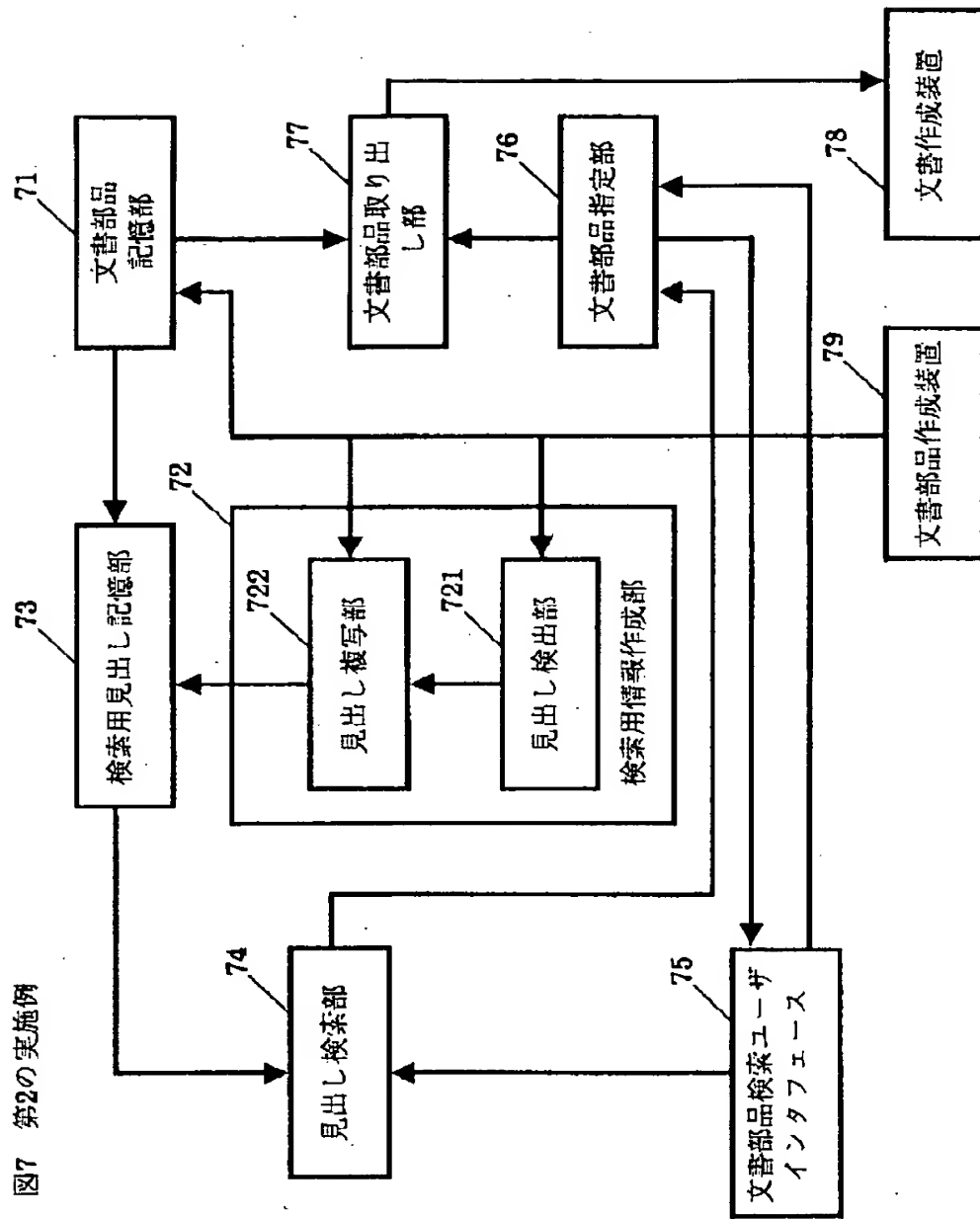


【図18】

図18

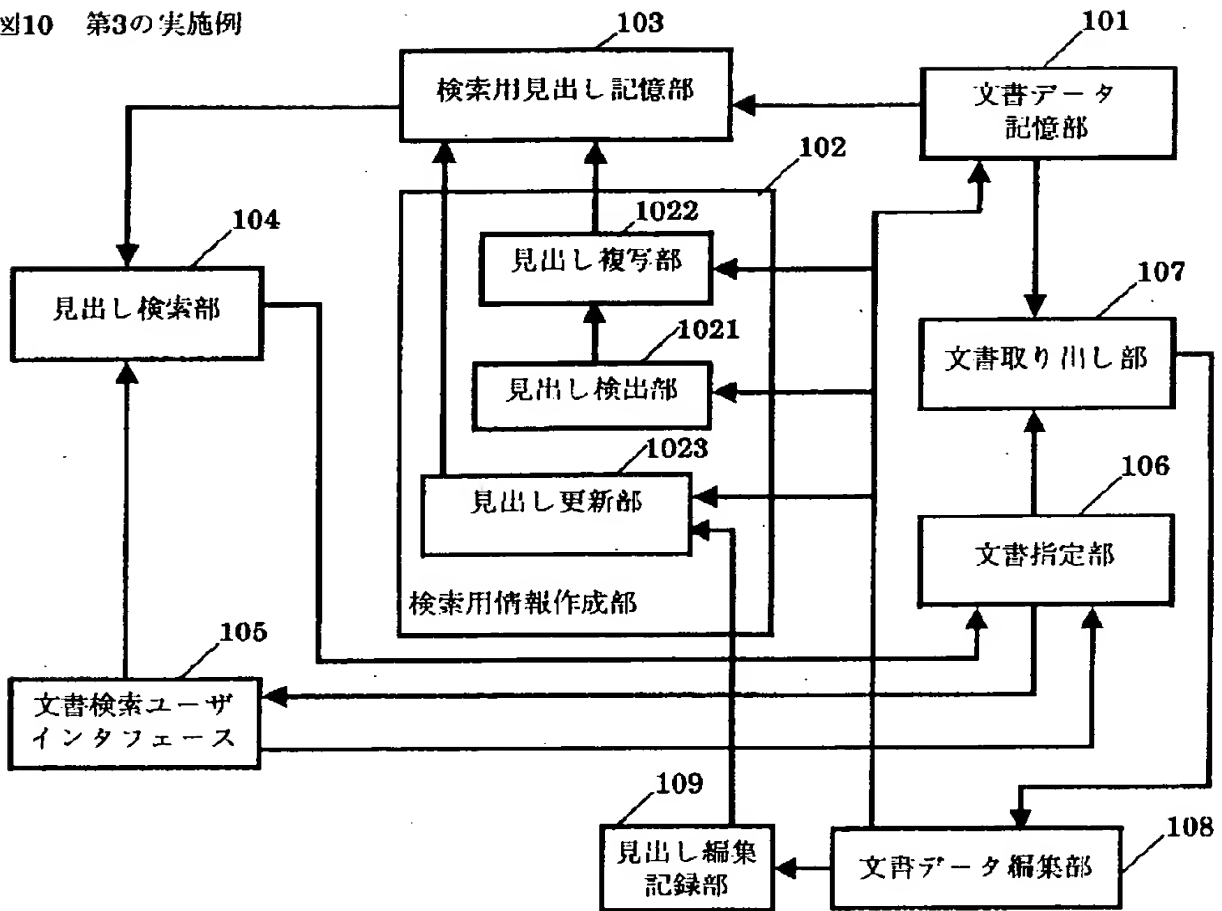


【図7】



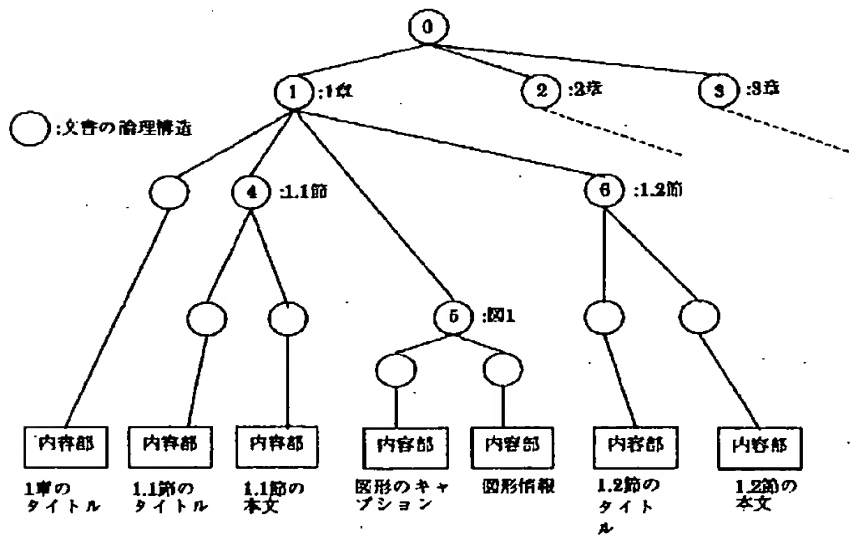
【図10】

図10 第3の実施例

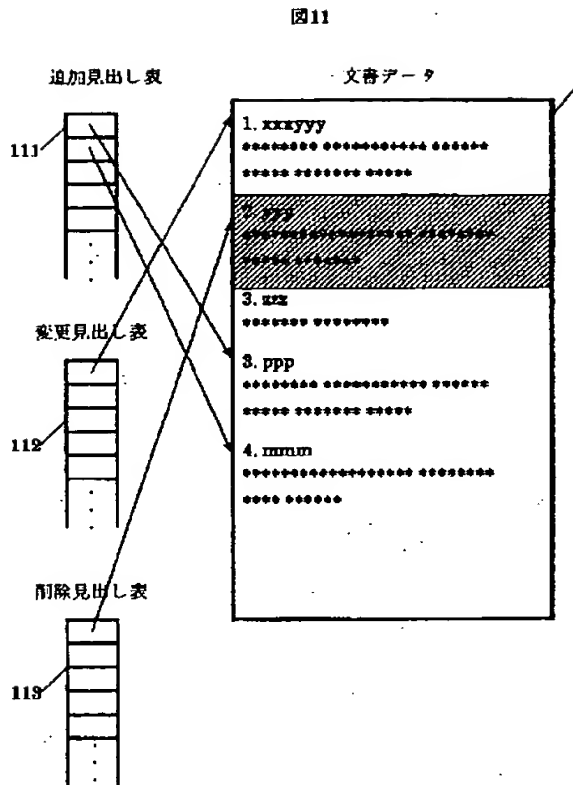


【図19】

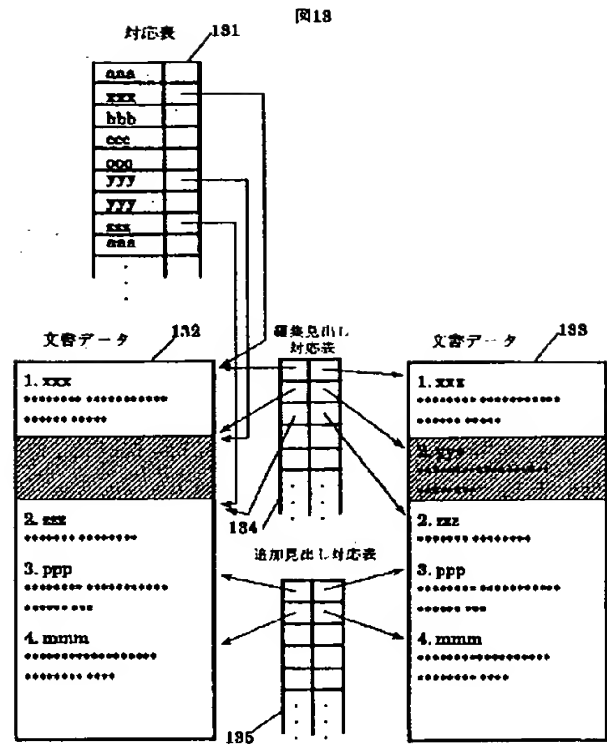
図19



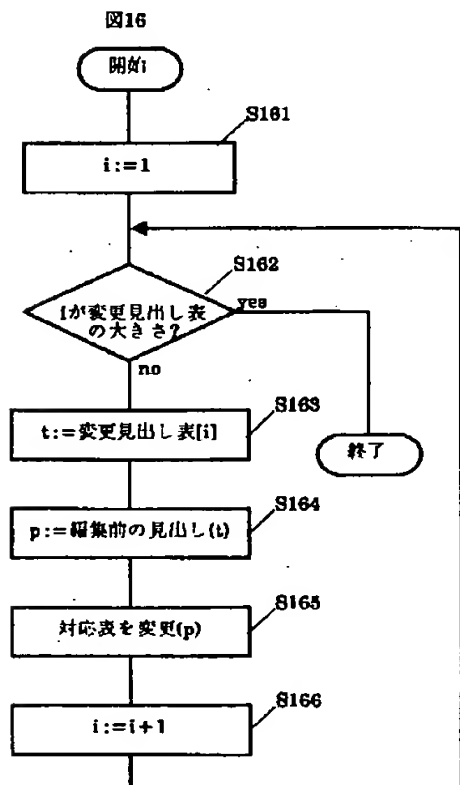
【図11】



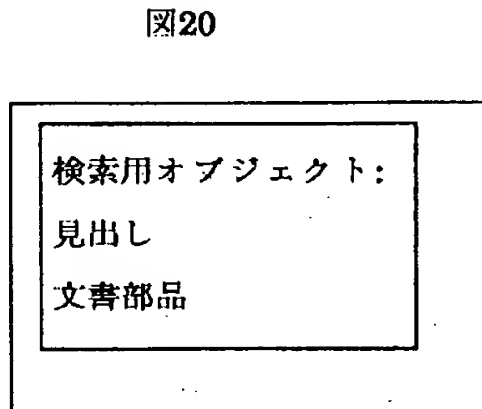
【図13】



【図16】

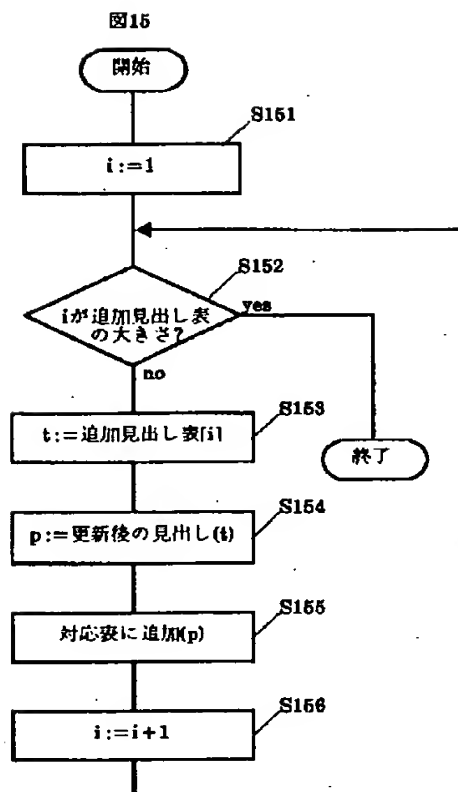


【図20】





【图 1-5】



【图 2-3】

42

データベースの技術動向  
富士太郎  
朝日新聞研究所

**Abstract:** オフィスにおける文書システムは、.....

## 1 始めに

## ● 文書システム

データベースシステムの、

[illegible]

## 1.1 RDBについて

## 1.1 RDBについて

### 3 オフィスの動向

ワークステーション上の.....

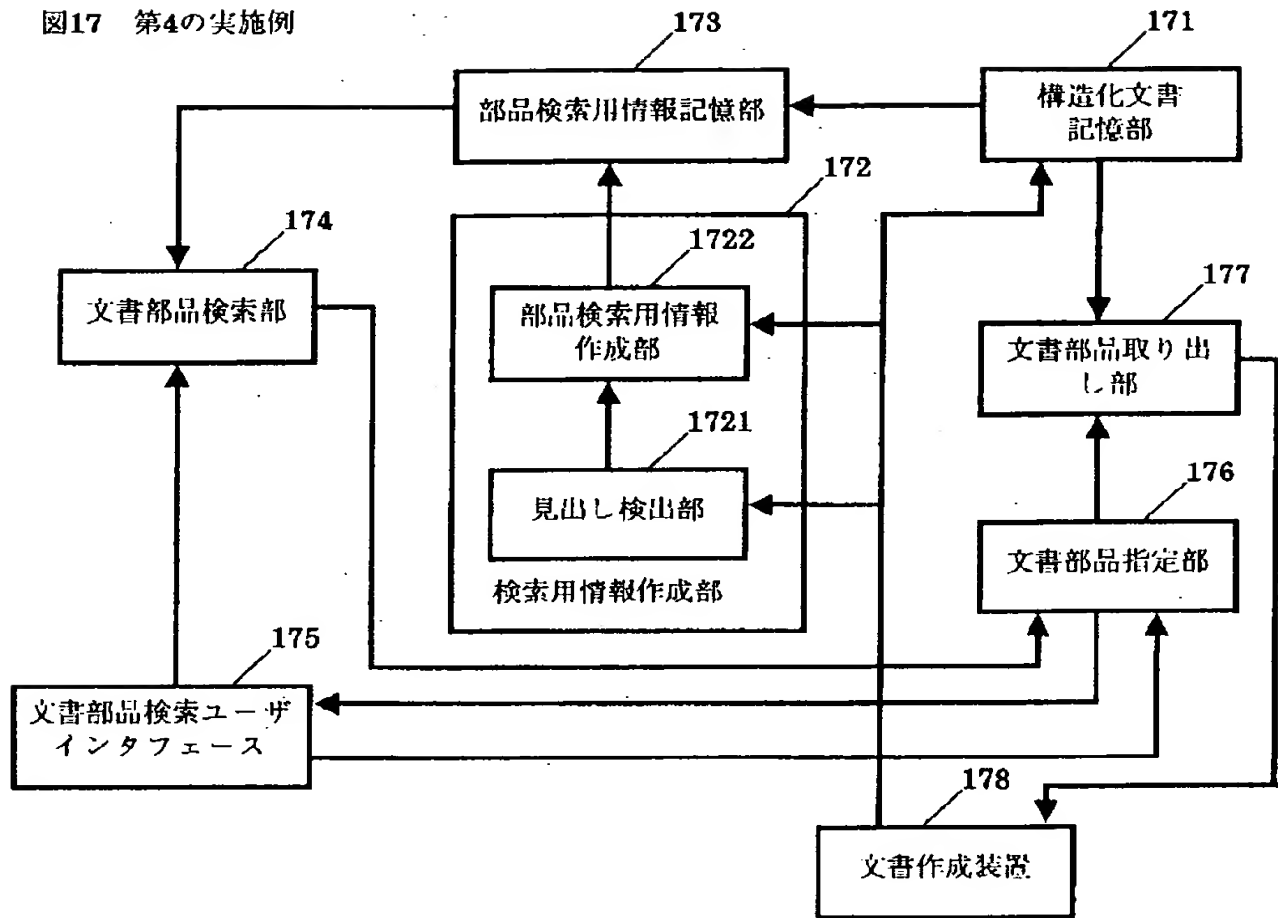
TABLE 1. The number of cases of the disease in the United States, 1960-1969, by age group and sex.

\*\*\*\*\*

\*\*\*\*\*

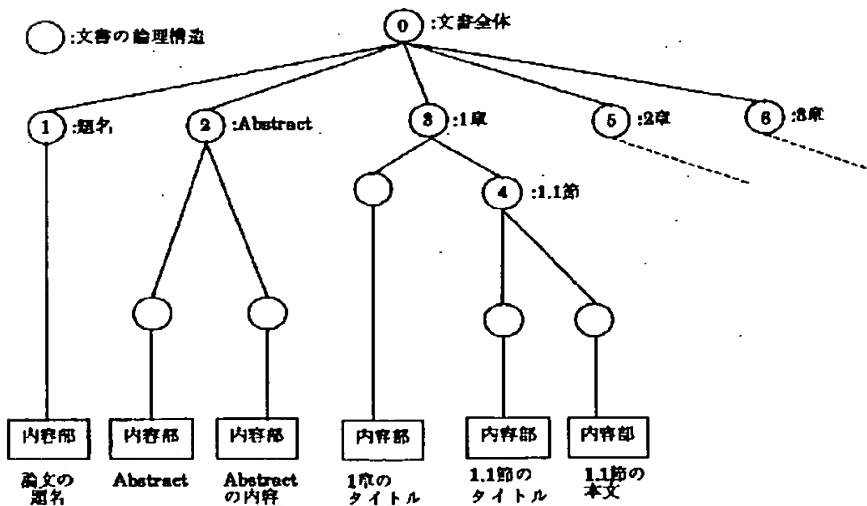
【図17】

図17 第4の実施例



【図24】

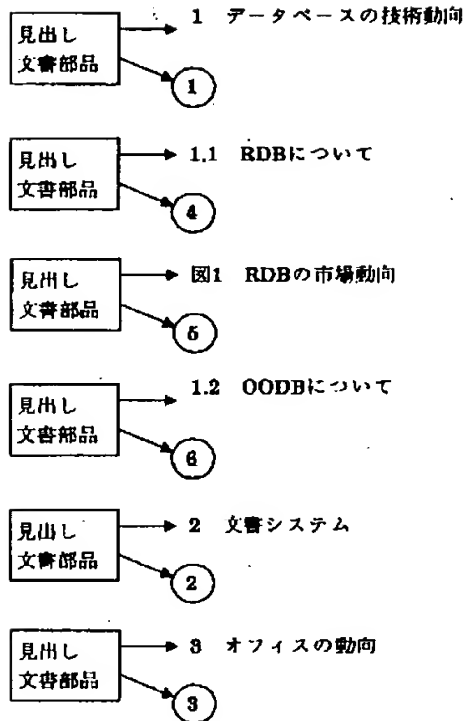
図24





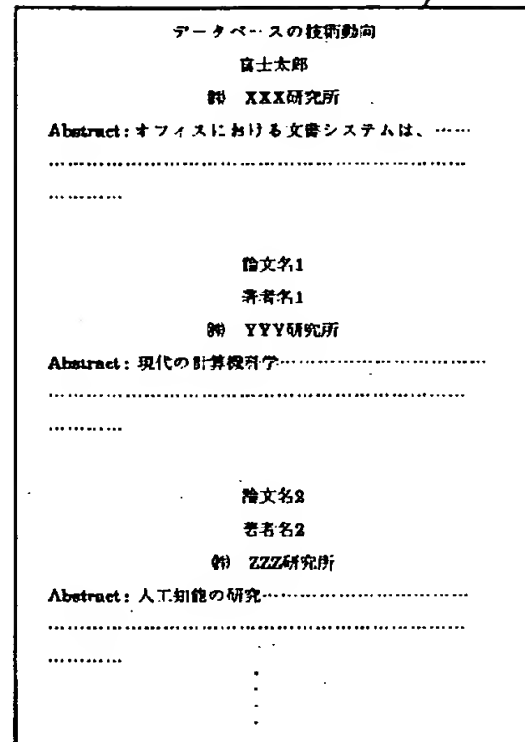
【図21】

図21



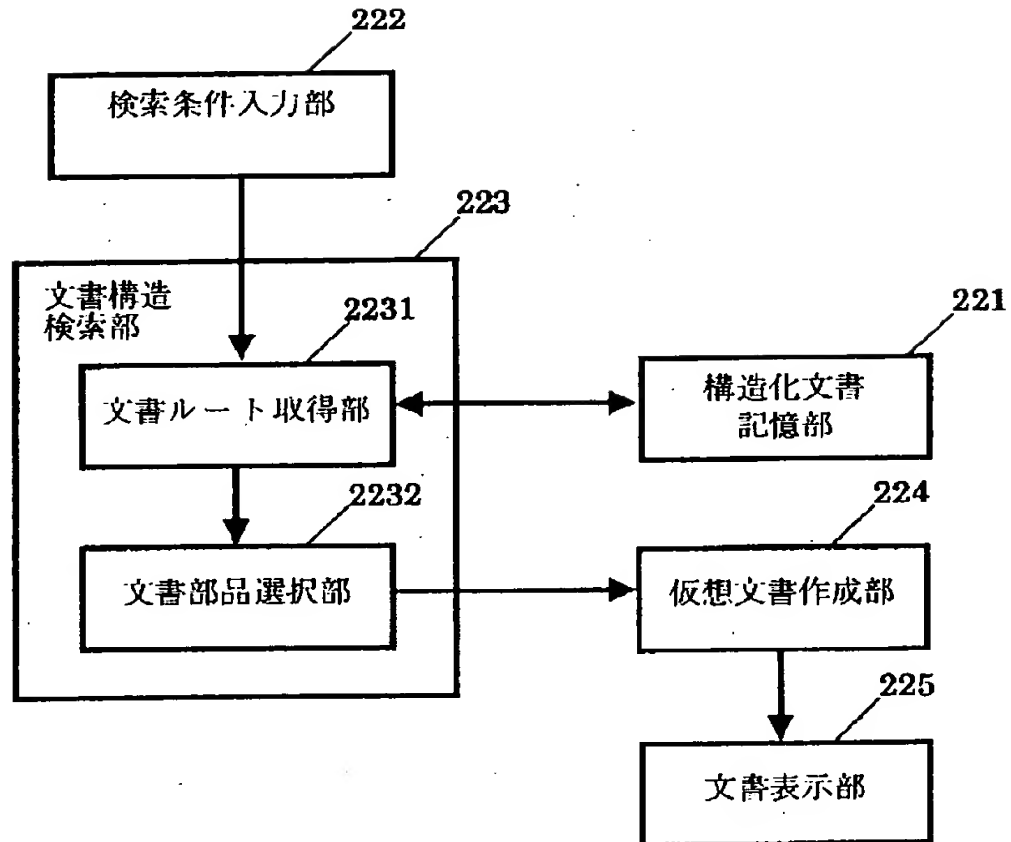
【図27】

図27



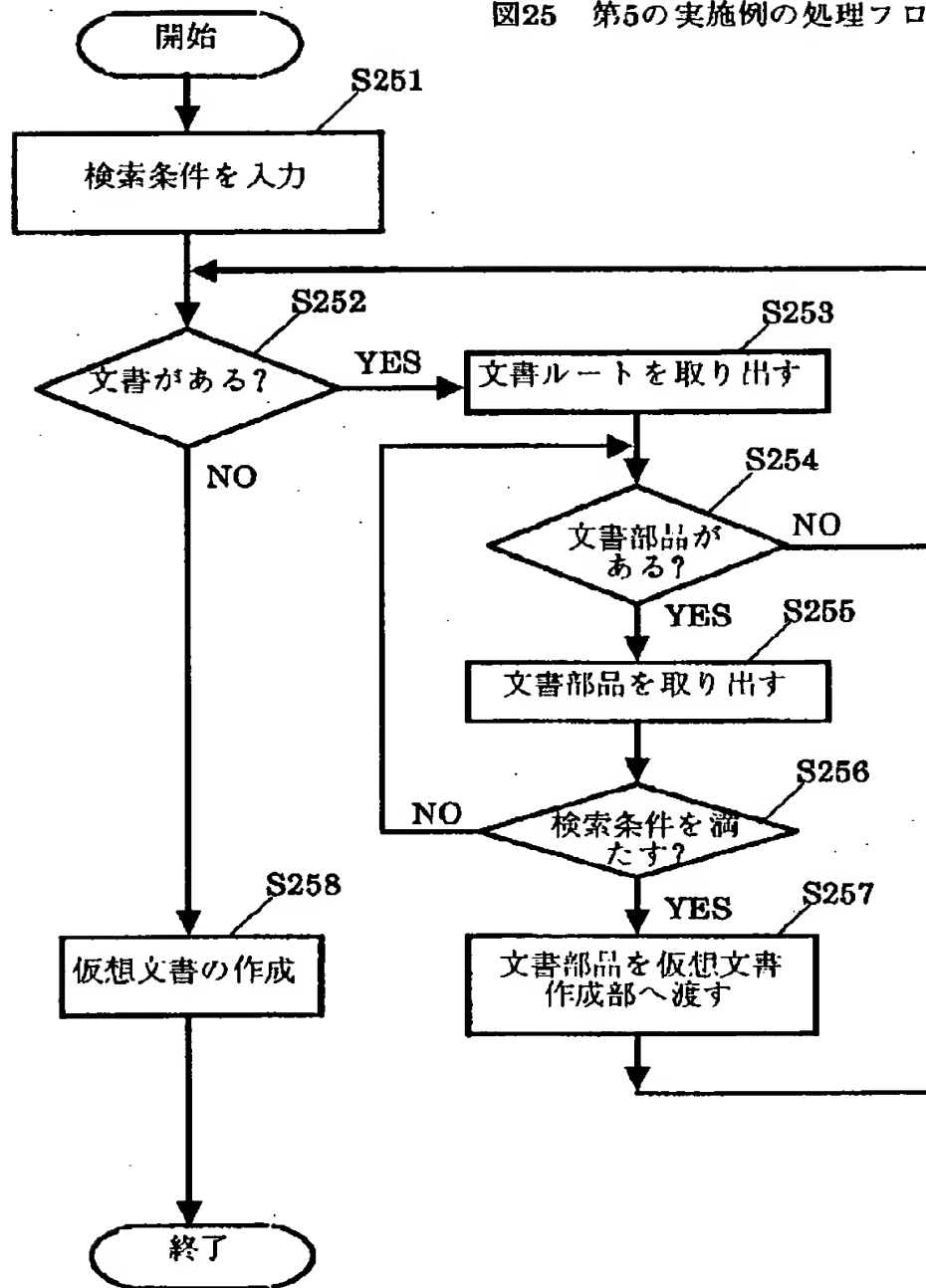
【図22】

図22 第5の実施例



【図25】

図25 第5の実施例の処理フロー



【図26】

図26

